PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/54, 9/12, 15/70, 15/85, C12Q 1/68 — A1

(11) International Publication Number: **WO 00/42199**

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/US00/00896

(22) International Filing Date: 14 January 2000 (14.01.00)

(30) Priority Data:
60/116,099 15 January 1999 (15.01.99) US

(71) Applicant *(for all designated States except US)*: MOLECULAR BIOLOGY RESOURCES [US/US]; 5520 West Burleigh Street, Milwaukee, WI 53210 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: SWAMINATHAN, Neela [US/US]; 3026 Boshard Drive, Madison, WI 53711 (US).

(74) Agent: MERKEL, William, K.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606-6402 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: BIOLOGICALLY ACTIVE REVERSE TRANSCRIPTASES

(57) Abstract

The invention provides modified reverse transcriptase polypeptides (Types I, II, and III), along with polynucleotides encoding such polypeptides, vectors containing such polynucleotides and host cells transformed with those polynucleotides. The modified RTs typically exhibit improved stability and/or improved solubility, relative to naturally occurring reverse transcriptases. The modified RTs are also found in a variety of forms, such as monomers as well as both homo- and hetero-multimers. The modified RTs may be used in any one or more of the methods known to benefit from reverse transcriptase activity, such as cDNA synthesis, and amplification techniques such as PCR and RAMP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

BIOLOGICALLY ACTIVE REVERSE TRANSCRIPTASES

FIELD OF THE INVENTION

In general, the invention relates to the field of molecular biology. In particular, the invention relates to reverse transcriptases.

BACKGROUND OF THE INVENTION

The defining activity of a reverse transcriptase (RT) is its ability to synthesize a cDNA strand using an RNA template. This activity has been exploited in a wide variety of techniques fundamental to progress in the academic and commercial arenas. For example, reverse transcription is useful in the production of cDNA molecules and libraries, sequence-specific probes having a variety of labels, sequencing techniques, and any of several amplification techniques. These amplification techniques include Reverse Transcription-Polymerase Chain Reaction (RT-PCR; *Myers et al., Biochemistry 30:*7661-7666 (1991) and U.S. Patent Nos. 5,310,652 and 5,407,800), Nucleic Acid Sequence-Based Amplification (NASBA; *Kievits et al., J. Virol. Methods 35:*273-286 (1991) and U.S. Patent Nos. 5,130,238 and 5,409,818), Self-Sustained Sequence Replication (3SR; *Guatelli et al., Proc. Natl. Acad. Sci. (USA) 87:*1874-1878, 1990) and Rapid Amplification (RAMP; PCT/US97/04170). Other amplification techniques take advantage, at least in part, of the DNA-dependent DNA polymerase activity of some RTs. Amplification techniques falling within this category include, *e.g.*, the Polymerase Chain Reaction (*i.e.*, PCR; Saiki *et al.*, Science 239:487-491 (1989) and U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159), the Inverse Polymerase Chain Reaction, the Multiplex Polymerase Chain Reaction, Strand Displacement Amplification (*i.e.*, SDA; Walker *et al.*, Proc. Natl. Acad. Sci. (USA) 89:392-396 (1992), Walker *et al.*, Nucl. Acids Res. 20(7):1691-1696 (1992), and U.S. Patent Nos. 5,270,184, and 5,455,166), and the Multiplex Strand Displacement Amplification (U.S. Patent No. 5,422,252 and 5,470,723).

Reverse transcriptases are found in a variety of retroviruses, or RNA tumor viruses. Techniques for producing RT from these native sources involve isolation of virus particles which contain about thirty RT molecules per virion. The RT is released from the virions by lysis of the virion coat. Released native RTs may then be purified using conventional techniques. However, the procedure involved in the production of these viruses is labor-

intensive and costly (1,000 infected chicks produce 10-20 grams of virus, which is approximately 25,000-40,000 units/gram of virus). Additional problems with RT production from natural sources are the high natural mutation rates which, in part, result in restricted host ranges such as specific strains of chickens.

An alternative source of RTs is recombinant production, which in turn is dependent on an understanding of RT expression by the various retroviruses. In general terms, retroviruses bind to receptors on susceptible cells and insert the retroviral core particle into the cytoplasm of the host. Two major events occur in the life cycle of retroviruses. First, the single-stranded RNA genome is converted to double-stranded DNA by reverse transcriptase. Second, this DNA copy is inserted into the genome of the host cell (Varmus, et al., *In Mobile DNA* (ed. Berg, et al.,) pp 53-108, (1989), Washington D.C.: AM. Soc. Microbiol. 972 pp; Brown, Curr. Top. Microbiol. Immunol. 157: 19-48 (1990); Goff, Cancer Cells 2: 172-178 (1990a); Goff, J. Acquired Immune Defic. Syndr. 3:817-31 (1990b); Boeke, et al., Curr. Opin. Cell. Biol. 3: 502-507 (1991), an event typically mediated by a virally encoded integrase activity. Following integration, this proviral DNA can be transcribed by the host RNA polymerase to make viral RNA which is then transported back to the cytoplasm for synthesis of various viral proteins. Virus assembly takes place in the cytoplasm followed by release of budded viruses from the cell for another round of infection (Whitcomb, et al., Ann. Rev. Cell Biol. 8: 275-306 (1992)). Any defect in the reverse transcription or integrase functions will result in a defective virus that cannot replicate. As an example, Avian Myeloblastosis Virus (*i.e.*, AMV) is a defective virus that requires a helper virus such as Myeloblastosis-Associated Virus (*i.e.*, MAV) for viral propagation.

Integrase ensures a stable association of viral and host DNAs. Integration is site-specific with respect to the viral DNA but is essentially random with respect to the host. This observation indicates that there is a DNA binding region in the integrase domain that is necessary for the binding of viral and host DNAs, in a manner independent of host sequence, during the integration process.

Although encoded by the cognate genes, the integrase domain is not found within mature MMLV-RT (*i.e.*, Moloney- Murine Leukemia Virus Reverse Transcriptase, a Type I RT) or mature HIV-RT (*i.e.*, Human Immunodeficiency Virus Reverse Transcriptase, a Type II RT). However, the integrase domain is found as an integral part of the mature avian RT (a Type III RT). The presence of this integral integrase domain, along with

thermostability, are two features of avian RTs that distinguish this class of RT from other RTs. Investigations of the integrase domain of avian RTs have revealed that it functions in DNA binding and in polymerization, or multimerization.

Some evidence for a DNA binding function comes from alignment of the deduced amino acid sequences of retroviral integrases. Three potential functional domains have been identified. An N-terminal region is characterized by an HHCC (Histidine, Cysteine) zinc finger-like domain which stabilizes the structure of the integrase (approximately, amino acids 579-629 of SEQ ID NO:2). The central region of these integrases contains a catalytic domain which shares homology with bacterial transposases involved in the breaking and joining of nucleic acid molecules (approximately, amino acids 630-807 of SEQ ID NO:2). This region has acidic amino acid residues which have been proposed to be involved in the binding of required metals (Mg^{++} or Mn^{++}). Khan *et al.*, Nucl. Acids Res. 19:851-860 (1991), reported DNA binding activity in this central region. The C-terminal region of these integrases is not conserved at the sequence level and its function is unknown (approximately, amino acids 808-858 of SEQ ID NO:2). However, deletion analyses indicate that this region contains strong sequence-independent DNA binding activity as well.

The integrase polypeptide functions as a multimer, or polymer. The N-terminal zinc finger-like domain and the C-terminal deletion derivative have less tendency to dimerize. Hickman *et al.*, J. Biol. Chem. 269:29,279-29,287 (1994). Sedimentation analyses suggest that integrase occurs as a mixture of monomers, dimers and tetramers.

The genome of the retroviruses codes for several genes, namely *gag*, *pol*, *env*, and the cellular oncogenes, *tat*, *ars/trs*, *nef*, *rev* etc. The *pol* gene codes for a polypeptide with reverse transcriptase (RT) activity. The RT enzyme has several activities, such as RNA-dependent DNA polymerase, DNA-dependent DNA polymerase, ribonuclease (RNase H), integrase, endonuclease and, possibly, protease activities. In the laboratory, reverse transcriptase is mainly used for its RNA-dependent DNA polymerase activity, which elongates an oligonucleotide primer, such as a tRNA, annealed to a template RNA or DNA strand to synthesize a DNA strand that is complementary to the template strand (cDNA) (Copeland, et al., J. of Virology 36: 115-119 (1980); Berger, et. al., Biochemistry 22: 2365-2372) (1983)).

Generally, there are three types of RT. Moloney-Murine Leukemia Virus (MMLV) is a monomeric RT, while HIV-RT and avian RTs are heterodimers. The HIV-RT

heterodimer consists of a 66 kDa β polypeptide and a 51 kDa α polypeptide. The avian RT heterodimer consists of a larger 95 kDa β polypeptide and a 63 kDa α polypeptide. The α polypeptides from HIV-RT and the avian RTs differ in that the HIV-RT α polypeptide lacks RNase H activity. The β polypeptide of HIV-RT and the β polypeptide of avian RTs differ in that the HIV-RT β polypeptide lacks the integrase activity of avian RT β polypeptides.

AMV-RT occurs in nature in multiple molecular forms, such as monomers, homodimers and heterodimers. However, the major active native form is a heterodimer of two structurally related polypeptide chains, an α subunit of 63 kDa and a β subunit of 95 kDa. These mature subunits are the products of post-translational processing of a precursor protein of 180 kDa (Gag+Pol). The 180 kDa protein is cleaved to a 95 kDa β subunit. The β subunit may be further cleaved to a 63 kDa α subunit and a 32 kDa endonuclease subunit. The α and β subunits have identical N-termini. (Roth, *et al.*,J. Biol. Chem. 260:9326-9335 (1985); Gerard, *et al.*,DNA 5: 271-279 (1986)).

Beyond a difference in form (monomer v. heterodimer), the avian RTs differ from MMLV-RT in other ways. In contrast to MMLV-RT, the avian reverse transcriptases exhibit high processivity and yield, as well as biological activity (*e.g.*, polynucleotide polymerase activity) over a wider range of temperatures extending up to at least 70°C. This ability to polymerize at higher temperatures is useful when working with RNA templates that have secondary structures. Additionally, this temperature stability has been exploited in amplification technologies such as NASBA and RAMP. Non-avian RTs, including those RTs having RNase H activity, have relatively low processivity and yield. For example, it has been estimated that approximately 50 times more MMLV RT is required than AMV-RT for cDNA synthesis.

In addition to Avian Myeloblastosis Virus, the avian retroviruses include Avian Sarcoma Leukosis Virus (ASLV), Rous Sarcoma Virus (RSV), Avian Sarcoma virus (ASV), Avian Tumor Virus (ATV) and their helper viruses such as MAV, Avian Sarcoma helper virus UR2AVRT, Rous-Associated Virus (RAV), and others. The homology among the avian reverse transcriptases at the DNA level is between 90-98% and, at the amino acid level, the homology is 95-100%.

Although the nucleotide sequences of many avian viruses are known (Schwartz *et al.*, Cell 32:853-869 (1983); see also Genbank Accession Nos. M24159, M37980, J02342,

J02021, and J02343), cloning and expression of an active and stable RT in commercially useful amounts has not been achieved.

When the DNA sequence of the *pol* gene of AMV and MAV were compared, approximately 111 bp from the 3' end of MAV was found to be replaced by host DNA sequences in AMV. Kan *et al.*, Virology 145: 323-329 (1985). The rest of the DNA coding for the RNA- and DNA-dependent DNA polymerase and RNase H activities was intact. This deletion involved the coding region for the integrase domain of the β polypeptide, which causes AMV to be defective in the propagation of the virus, thereby creating a requirement for helper virus MAV to produce infectious progeny virus. Hence, the integrase domain is critical for producing infectious particles. Nevertheless, both the avian retroviruses and their helper viruses encode reverse transcriptases having RNA- and DNA-dependent polymerase and RNase H activities.

AMV Reverse Transcriptase (*i.e.*, AMV-RT) has been characterized and conditions for the synthesis of full-length cDNA products have been investigated. Berger *et al.*, Biochemistry 22:2365-2372 (1983). However, the length and yield of cDNA produced by AMV-RT have reportedly been limited by either a nuclease integral to AMV-RT or associated contaminants. See, U.S. Patent No. 5,017,492. In efforts to maximize cDNA length and yield, attention has turned to MMLV-RT. MMLV-RT is a reverse transcriptase that is relatively thermosensitive and exhibits relatively low reverse transcriptase activity. Efforts to improve the stability, and hence activity, of MMLV-RT reportedly met with some success in the form of C-terminal truncations of MMLV-RT. U.S. Patent No. 5,017,492; see also U.S. Patent Nos. 5,244,797, 5,405,776, and 5,668,005. Beyond these modifications, the '492 Patent reports that some C-terminal amino acid changes enhanced MMLV-RT activity, albeit at the cost of a reduction in processivity. Notwithstanding these improvements, MMLV-RT is relatively thermosensitive and inefficient in catalyzing cDNA synthesis.

The avian RTs are structurally distinct from MMLV-RT. At the primary structure level, avian RT, *e.g.*, AMV-RT, shares no more than 28% amino acid sequence similarity to MMLV-RT (no more than 50% similarity at the polynucleotide level). Moreover, the native AMV-RT is a heterodimer composed of a 63 kDa alpha peptide and a 95 kDa beta peptide while MMLV-RT is an 80 kDa monomer. Not surprisingly, these enzymes differ in their thermostability. The thermophilic AMV-RT is active over a broad temperature range extending, at least, to 70°C. Consequently, these avian RTs can often copy RNA

templates capable of forming relatively strong secondary structures. In contrast, MMLV-RT is a mesophilic enzyme. Also, relative to AMV-RT, approximately 50-fold more MMLV-RT is required for cDNA synthesis. Furthermore, AMV-RT and MMLV-RT differ in other properties such as processivity, metal co-factor requirements, error rate (*i.e.*, rate of incorrect nucleotide incorporation), and tRNA primer preferences. These drawbacks in using MMLV-RT, in turn, increase the cost of effectively using MMLV-RT. Therefore, a need continues to exist in the art for a reverse transcriptase that can be produced economically and that exhibits one or more improvements in terms of processivity, stability, solubility, and thermal range, leading to increased lengths and yields of polynucleotide products, while minimizing the cost of the reverse transcriptase.

SUMMARY OF THE INVENTION

The present invention relates to the discovery that reverse transcriptase polypeptides which have been modified, *e.g.*, by altering existing integrase domains or by adding integrase domains that is modified themselves, are characterized by one or more improved properties, which include increased activity, stability, and solubility, as well as increased ease and versatility in producing such polypeptides. The reverse transcriptase polypeptides of the invention may be derived from any source, including, but not limited to, Moloney-Murine Leukemia Virus (a Type I reverse transcriptase or RT), HIV (Type II RTs), and avian retroviruses (Type III RTs). One aspect of the invention is drawn to RT polypeptides that are truncated internally and/or at their C-termini, yet retain RNA-dependent DNA polymerase activity, the defining characteristic of reverse transcriptases. The truncated polypeptides may also have, and preferably do have, DNA-dependent DNA polymerase activity. Preferred polypeptides according to the invention exhibit RNase H activity. For those truncated polypeptides corresponding to full-length reverse transcriptases having an integral integrase activity (*e.g.*, avian retroviral RTs or modified Type I and Type II RTs that retain an integrase domain, unlike natural forms of these RTs), the truncation preferably extends into the integrase domain, effectively eliminating integrase activity from the truncated polypeptide. Such truncated polypeptides exhibit improvements in one or more of the following properties compared to their full-length counterparts: RNA-dependent DNA polymerase activity, expression levels, stability, and solubility. These improvements result in more cost-effective RTs for use in a wide variety of DNA synthesis, amplification and sequencing technologies.

The invention also provides a chimeric RT polypeptide resulting from the effective addition of a protein domain to the C-terminus of the truncated RT, resulting in a non-native chimeric polypeptide (*i.e.*, a polypeptide not found in nature). These protein domains provide a DNA binding capability, a metal binding capability, a structure stabilizing capacity, or a polymerization (*i.e.*, multimerization) capability, and preferably several capabilities. With these added, or enhanced, capabilities, the chimeric polypeptides of the invention exhibit improvements in RNA-dependent DNA polymerase activity, protein expression levels, protein stability, and/or protein solubility, with chimeric polypeptides of the invention frequently showing improvement in all four properties. Preferred protein domains include a plurality of histidine residues (*i.e.*, His tags), and either the N-terminal domain (providing a DNA binding capacity, preferably resulting from a zinc finger domain) or the C-terminal domain (providing a polymerization domain) of the integrase region of a native RT.

More specifically, the invention provides reverse transcriptase polypeptide fragments (*i.e.*, portions of full-length RT polypeptides), modified reverse transcriptase polypeptides, and analogs and variants thereof. Preferably, the polypeptides of the invention are thermostable avian RTs that have improved RNA- and DNA-dependent DNA polymerase activities, resulting in increased lengths and yields of synthesized polynucleotide products. Typically, the polypeptides of the invention lack the catalytic activity of the integrase domain provided by the C-terminal region of the full-length polypeptides (*e.g.*, nucleotides 1719-2571 of SEQ ID NO:1 (Type III), nucleotides 2464-3012 of SEQ ID NO:40 (Type I), and nucleotides 1840-2708 of SEQ ID NO:42 (Type II)). The absence of catalytic activity provided by the integrase domain is expected to result in polypeptides that are more soluble and expressed at higher levels, hence, such polypeptides are more amenable to economical purification in commercially useful quantities. In addition to this benefit, the chimeric polypeptides of the invention are expected to facilitate nucleic acid binding or polymerization (homo-polymerization or hetero-polymerization), and preferably both activities, which contribute to the improved performance of the polypeptides. The improved RT performance, in turn, translates into improvements in the many techniques dependent on RT activity, such as cDNA production and cDNA library preparation as well as a variety of polynucleotide amplification and sequencing technologies. These amplification techniques include RT-PCR, NASBA, 3SR, and RAMP. The improved DNA-dependent DNA polymerase activities of the polypeptides of the invention are useful in, *e.g.*, PCR, the Inverse Polymerase Chain Reaction, the Multiplex Polymerase Chain Reaction, SDA, and Multiplex SDA. The sequencing technologies include the many variations on the Sanger dideoxy sequencing technique.

One aspect of the invention is an isolated polynucleotide encoding a polypeptide according to the invention. In general terms, the invention comprehends polynucleotides encoding polypeptides having RT activities, those polynucleotides typically lacking approximately 200-1,122 bp of the 3′ ends of the corresponding native RT genes. For example, a full-length avian RT gene (*i.e.*, MAV *pol*) is 2,692 bp (SEQ ID NO:1) and the invention contemplates MAV-derived polynucleotides of approximately 1,570-2,492 bp in length. More generally, the polynucleotides of the invention may result from truncations to RT-encoding polynucleotides derived from any source, including: AMV, MAV, RSV, ASLV, ATV, MMLV and HIV. In particular, the invention contemplates an isolated polynucleotide encoding a polypeptide having RNA-dependent DNA polymerase activity, the polypeptide consisting of any one of the following sequences: an amino acid sequence

beginning at amino acid 1 and terminating at any one of amino acids 428 to 857 of SEQ ID NO:2; an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 428 to 1,054 of SEQ ID NO:39; an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 548 to 1,198 of SEQ ID NO:41; an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 428 to 901 of SEQ ID NO:43; and variants, analogs and fragments of any of the above-described polypeptides having RNA-dependent DNA polymerase activity, the aforementioned polypeptides (*i.e.*, polypeptides and variants, analogs, and fragments thereof) optionally having an N-terminal methionine. An exemplary polynucleotide has a sequence set forth in any one of SEQ ID NOs 1, 7, 9, 38, 40, and 42. The polynucleotides preferably comprise a start codon specifying methionine at the 5′ end. Other truncated polynucleotides of the invention have internal deletions, preferably removing at least part of an integrase domain. For example, polynucleotides according to the invention comprise the sequence set forth in SEQ ID NO:40, with part or all of nucleotides 2464-3012 deleted, or comprise the sequence set forth in SEQ ID NO:42, with part or all of nucleotides 1840-2708 deleted, or comprise the sequence set forth in SEQ ID NO:1, with part or all of nucleotides 1719-2571 deleted (e.g., deletion of nucleotides 1860-2310, 1920-2310, or 1980-2310 of SEQ ID NO:1). Such polynucleotides encode polypeptides that lack an effective integrase activity in that the polypeptides do not promote detectable polynucleotide integration.

Other polynucleotides according to the invention encode chimeric polypeptides, such polynucleotides comprising a polynucleotide encoding a polypeptide having RNA-dependent DNA polymerase activity and an adjacent polynucleotide encoding a terminal modification of that polypeptide, thereby encoding a chimeric polypeptide. Preferred polynucleotides encode a chimeric polypeptide having one or more amino acids attached to the C-terminus of a polypeptide having RNA-dependent DNA polymerase activity. Such polynucleotides may contain one of the above-described coding regions fused (in frame) at its 3′ end to a region encoding one or more amino acids. For example, the 3′ end of a coding region may be fused to one or more codons for a charged amino acid such as histidine, lysine, arginine, aspartate, or glutamate. Alternatively, the 3′ end of the coding region may be fused to a region encoding a polypeptide, preferably having four to fifty (*e.g.*, six) amino acids and preferably comprising a domain selected from the group consisting of a DNA binding domain, an RNA binding domain, a metal binding domain,

a polymerization domain, and a structure stabilizing domain. Examples of such domains include, but are not limited to, disulfide bond forming cysteine residues, a zinc finger domain, an acidic amino acid domain, and a basic amino acid domain, a bulky amino acid domain (*e.g.*, W or W-H, single-letter amino acid identifications), a PPG domain, a GPRP or a PRPG (*i.e.*, inverse GPRP) domain, a leucine zipper motif or domain, and an NS1 binding site, among others. Examples of suitable domains include, but are not limited to, the N terminal domain of the MAV-RT integrase region which provides a DNA binding domain and the C-terminal domain of the integrase region which provides a polymerization domain. Further, the polynucleotides encoding chimeric polypeptides having a plurality of C-terminal amino acids may encode the same amino acid a number of times. Such polynucleotides may encode basic (*e.g.*, Histidine) amino acids at the C-terminus. Also preferred are polynucleotides that have a stop codon (*e.g.*, TAA, TAG, or TGA) at the 3′ end of a coding region of a chimera according to the invention. An exemplary polynucleotide encoding a chimeric polypeptide has a sequence selected from the group consisting of a sequence set forth in any one of SEQ ID NOs 11-19.

Still other polynucleotides of the invention encode a chimeric polypeptide having one or more amino acids attached to the N-terminus of a polypeptide having RNA-dependent DNA polymerase activity. In addition, the invention contemplates polynucleotides that encode more than one modification, such as an N-terminal peptide addition and a C-terminal peptide addition or a C-terminal peptide addition coupled to an internal deletion of at least part of an integrase domain.

The invention also provides a vector comprising any of the aforementioned polynucleotides. A preferred vector comprises a polynucleotide operably linked to a promoter.

Another aspect of the invention is directed to a host cell transformed with a polynucleotide of the invention, such as prokaryotic (*e.g.*, *Escherichia coli*) or eukaryotic cells (*e.g.*, insect cells). In a related aspect, the invention comprehends a method of transforming host cells comprising the following steps: introducing a vector according to the invention into a host cell; incubating the host cells; and identifying host cells containing the vector, thereby identifying a transformed host cell.

Still another aspect of the invention is a method of producing an isolated reverse transcriptase polypeptide comprising the step of transforming a host cell with a vector as described above, incubating the host cell under conditions suitable for expression of a

polypeptide, and recovering the polypeptide, thereby producing an isolated reverse transcriptase polypeptide according to the invention.

In another aspect, the invention provides the polypeptides encoded by the polynucleotides described above. These polypeptides include polypeptide fragments (*e.g.*, β RT fragments containing part, but not all, of the C-terminal integrase domain) and chimeric polypeptides, as described above, as well as variants and analogs thereof. In general terms, the invention contemplates all types of reverse transcriptase fragments and chimeras (and variants and analogs thereof) including, but not limited to, the three classes of RTs exemplified by MMLV-RT, HIV-1-RT, and avian RTs. Exemplary chimeric polypeptides contain an N-terminal methionine or a C-terminal peptide providing useful functions (*e.g.*, expression enhancement, nucleic acid binding domains, metal binding domain, structure stabilizing domains, or polymer-forming domains). Other chimeric polypeptides according to the invention may result from modification of RTs derived from, *e.g.*, the following sources of Types I-III: ASLV, ATV, MMLV, HIV-1, and HIV-2. A preferred addition to an RT is a C-terminal peptide comprising a plurality of amino acids such as basic amino acids, a nucleic acid binding domain, a metal binding domain, or a polymerization domain. Preferably, the C-terminal peptide provides more than one functionally significant domain. Also preferred is one or more C-terminal cysteine residues, which, at a minimum, provide a capacity to induce polypeptide homo-, or hetero-, polymerization, such as dimerization. Typical polypeptides of the invention are relatively soluble and are capable of being expressed at high levels, resulting in relatively high levels of RT activity expected to facilitate economical purification.

Yet another aspect of the invention is an improvement in a method for copying a target nucleic acid by extending a target nucleic acid-bound primer, the improvement comprising: contacting the target nucleic acid and primer with a polypeptide according to the present invention. The method preferably produces one or more copies of the target nucleic acid and the polypeptide may be a polymer. Any method for copying a target nucleic acid using a polymerase is comprehended by the invention, including, but not limited to, cDNA synthesis, Polymerase Chain Reaction, Polymerase Chain Reaction-Reverse Transcription, Inverse Polymerase Chain Reaction, Multiplex Polymerase Chain Reaction, Strand Displacement Amplification, Multiplex Strand Displacement Amplification, Nucleic Acid Sequence-Based Amplification, Sequence-Specific Strand Replication and Rapid Amplification.

Another aspect of the invention is directed to improved methods for sequencing a target nucleic acid by extending a target nucleic acid-bound primer, the improvement comprising: contacting the target nucleic acid and primer with a polypeptide according to the present invention.

Yet another aspect of the invention is a kit for copying a target nucleic acid comprising one or more nucleotides and a polypeptide according to the invention. Preferred polypeptides include those polypeptides encoded by a polynucleotide having a sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42 and polynucleotide derivatives thereof encoding C-terminal amino acids or polypeptides at their 3′ ends.

Numerous other aspects and advantages of the present invention will be apparent upon consideration of the following drawing and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 photographically depicts Western blot analysis of RT expression products of insect cells.

Fig. 2 illustrates recombinant RT fractionated on an 8% SDS-PAGE gel and stained with Coomassie Blue.

Fig. 3 presents an autoradiograph of gel-fractionated cDNAs produced by an RT polypeptide according to the invention.

Fig. 4 graphically presents temperature profiles for cDNA production using native and recombinant RTs (Fig. 4A), temperature profiles of nRT and rRT catalyzing RT-PCR (Fig. 4B), temperature profiles for RT-mediated RAMP (Fig. 4C), pH profiles for nRT and rRT in RT assays (Figs. 4D and 4E), magnesium ion profile for nRT and rRT in RT assays (Fig. 4F), and other divalent cation profiles for nRT and rRT in RT assays (Fig. 4G).

Fig. 5 illustrates the relative DNA-dependent DNA polymerase activities of native RT and recombinant RT.

Fig. 6 shows a graphic comparison of the relative RNase activities of native RT and recombinant RT at 37°C (Fig. 6A); Fig. 6B shows a temperature profile for the RNase H activity of rRT.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides truncated reverse transcriptase polypeptides (*i.e.*, fragments), and analogs and variants thereof. Preferably, these polypeptides exhibit improved levels of RNA-dependent DNA polymerase activity, frequently extending over a wide range of temperatures up to 70°C and beyond. Also preferred are internally or terminally truncated polypeptides having sequences compatible with improved levels of expression. A preferred polypeptide according to the invention has a temperature optimum of 45°-55°C. Also preferred is a polypeptide consisting of an amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:39, SEQ ID NO:41, or SEQ ID NO:43. Some of these polypeptides correspond to C-terminal truncated forms of avian reverse transcriptases, such as the full-length Myelogenetic Avian Virus-Reverse Transcriptase (*i.e.*, MAV-RT). A preferred polypeptide of the invention lacks an effective integrase catalytic activity and is expressed at elevated levels, providing a source of soluble, and recoverable, polypeptide in active form. Exemplary integrase domains include a Type I domain (nucleotides 2464-3012 of SEQ ID NO:40), a Type II domain (nucleotides 1840-2708 of SEQ ID NO:42) and a Type III domain ((nucleotides 1734-2571 of SEQ ID NO:1), any of which may be modified by internal or terminal deletion(s) or by substitution or chemical modification. Because integrase and RT function sequentially in the viral life cycle, it is possible that RT and integrase act in a complex. Thus, without wishing to be bound by theory, the added functions of nucleic acid binding and polymerization provided by the integrase domain of avian RTs may result in increased processivity and superior performance of such RTs. Accordingly, non-native chimeric polypeptides of the invention further include the C-terminal addition of a polymerizing domain, such as a plurality of the same, or different, amino acids. Non-native chimeric polypeptides are herein defined as polypeptides not found in nature. Thus, if the parts of the chimera are found in nature, they are not found in the same relationship as exists in the non-native chimeric polypeptide. Preferred C-terminal amino acid additions are basic amino acids, such as histidine, lysine and arginine. These preferred C-terminal additions may promote polymerization by, *e.g.*, metal chelation; the basic amino acids also may provide or enhance the nucleic acid binding capacity of the polypeptide. A preferred number of C-terminal amino acid additions is 4-50, more preferably six amino acids. As

one alternative to a plurality of basic amino acids, one or more cysteine residues may be added to the C-terminus of the polypeptide. Other alternatives are C-terminal peptides of 4-50 amino acids having a polymerizing capacity or a DNA binding capacity, and preferably both capacities. In addition, to RNA-dependent DNA polymerase activity, the polypeptides may also have DNA-dependent DNA polymerase activities or RNase H activity.

The invention also comprehends polypeptide variants, which have substantially the same amino acid sequence as one of the polypeptides described above. "Substantially the same" means that the sequence of the polypeptide may be aligned with one of the sequences disclosed herein, using any of the approaches known in the art (*e.g.*, DNASIS, Hitachi Software Engineering America, Ltd., San Bruno, CA) such that the sequences are at least 90%, and preferably 95% or 98%, similar throughout the aligned region. For example, the invention contemplates the conservative substitution of asparagine for aspartate at any one or more of amino acid positions 450, 505, or 564 of SEQ ID NO:2 to produce variant MAV-RT polypeptides lacking RNase H activity; that same substitution at any one or more of amino acid positions 497, 552, or 603 of SEQ ID NO:43 produces variants of HIV-RT polypeptides lacking RNase H activity. Other residues which may be changed by conservative substitution to generate RNase H$^-$ variants of MAV-RT include amino acid positions 484, 549, and 572 of SEQ ID NO:2. More generally, the invention comprehends polypeptides having substantially the same amino acid sequences, regardless of whether the differences involve conservative substitutions or not. For example, the residues identified above may be changed in a non-conservative manner. In addition, other residues known to be involved in RNase H activity may be altered by substitution or deletion. These residues include, but are not limited to, amino acids at positions 441-578 of SEQ ID NO:2 (AMV-RT and MAV-RT; see also, RSV-RT); positions 427-1,055 of SEQ ID NO:39 (HIV-2-RT); positions 625-911 of SEQ ID NO:41 (MMLV-RT); and positions 427-902 of SEQ ID NO:43 (HIV-1-RT). The invention also comprehends polypeptide analogs, which are defined herein as polypeptides that either contain known equivalents for one or more of the conventional amino acids or have been derivatized in a manner understood in the art (*e.g.*, glycosylation, pegylation, phosphorylation), or both.

Another aspect of the invention is drawn to polynucleotides encoding the aforementioned polypeptides. A preferred polynucleotide consists of the sequence set forth as SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID

NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:38, SEQ ID NO:40, or SEQ ID NO:42. Also contemplated by the invention are polynucleotides substantially the same as the polynucleotides having one of the above-identified sequences. In the context of polynucleotides, "substantially the same" means that the polynucleotide has a sequence that is at least 90% homologous to one of the above-described polynucleotides.

Beyond the polynucleotides, the invention provides vectors containing at least one of these polynucleotides. Further, these vectors may be functional in prokaryotic cells, eukaryotic cells, or both cell types. A preferred vector is a Baculovirus vector such as pBacPak9 (Clontech Inc. Palo Alto, CA). The invention also provides prokaryotic and eukaryotic host cells transformed with the above-identified polynucleotides. A preferred host cell is an Sf9 insect cell transformed with a Baculovirus-based recombinant molecule of the invention. Other insect cell lines, such as SF21 HighFive may also be used.

In another aspect, the invention provides methods of using the polynucleotides to produce RTs according to the invention. In particular, the polynucleotides are transformed into a prokaryotic or eukaryotic host cell under conditions that allow expression of the encoded RT polypeptide and, following an incubation period, the polypeptide is isolated.

In yet another aspect of the invention, methods of using the RT polypeptides are provided. These methods realize the benefits of speed and yield from using highly active and thermostable RT polypeptides to copy target nucleic acids (*e.g.*, cDNA synthesis, cDNA library construction), amplify, or sequence a target nucleic acid. Suitable amplification methodologies include, but are not limited to, PCR, RT-PCR, Inverse PCR, Multiplex PCR, SDA, Multiplex SDA, NASBA, 3SR, and RAMP. Suitable sequencing methodologies include the original enzymatic sequencing technology disclosed by Sanger and co-workers, or any of the numerous variations of that technique that have been developed since that disclosure.

Various aspects of the invention are described in the following Examples, wherein Example 1 describes the cloning of a coding region encoding the full-length MAV-RT; Example 2 describes the sequencing of the full-length MAV *pol* gene encoding reverse transcriptase; Example 3 discloses the cloning of selected polynucleotides according to the invention; Example 4 details the large-scale purification of the expressed recombinant RT; Example 5 describes SDS-PAGE and Western blot analyses of expressed proteins; Example

6 discloses an assay for RNA-dependent DNA polymerase activity; Example 7 illustrates assays characterizing the native reverse transcriptase (nRT) and recombinant reverse transcriptase (rRT) in terms of optima for temperature, pH, $MgCl_2$, and other divalent cation concentrations; Example 8 discloses use of RTs in methods for copying and/or amplifying target nucleic acids; Example 9 describes a DNA-dependent DNA polymerase assay used to characterize nRT and rRT; Example 10 reports a comparison of the RNase H activities of nRT and rRT; and Example 11 describes the cloning and expression of additional polynucleotides according to the invention.

Example 1

The *pol* gene of MAV, encoding the full-length RT precursor polypeptide, was cloned from pMAV, a pBR322 derivative containing the *pol*, *gag* and partial *env* gene of MAV. Data derived from a partial restriction map of the insert fragment of pMAV is shown in Table I. Based on the map data, the *pol* coding region, along with some 5′ and 3′ non-coding sequences, was excised and ligated into several prokaryotic and eukaryotic vectors, as described below. Several recombinants were obtained from these vectors. Anti-RT monoclonal antibodies were used to analyze the expression of RT (*see* Example 5).

Table I

Feature	Relative Position (bp)
*Eco*RI	69
*Pst*I	200
Start codon (*pol*)	253
*Bgl*II	1988
*Kpn*I	2748
Stop codon (*pol*)	2943
*Xho*I	3013
*Pst*I	3155

All non coding 5′ (*i.e.*, upstream) nucleotides were removed to increase the expression of RT. Also, the open reading frame of the natural RT gene starts with an "ACT" (Thr), which is not a frequently used start codon in prokaryotes. The codon that is most frequently used is "ATG" (Met). "ATG" can serve as a start codon for efficient expression of RT in both prokaryotes and eukaryotes. Therefore, an "ATG" was added 5′ to the natural "ACT" start codon in order to allow efficient expression of the protein in prokaryotes and eukaryotes (**ATG ACT** GTT GCG CTA CAT CTG GCT ATT CCG CTC AAA TGG AAG CCA AAC CAC ACG CCT GTG TGG ATT TTC CAG TGG CCC, etc.; compare the sequences provided in SEQ ID NOs 2 and 3).

Construction of Prokaryotic Recombinant Vectors

pH contains a strong and tightly regulated lambda P_R promoter, a temperature sensitive λ cI repressor, an *E. coli* origin of replication, and Amp^r for selection. Because this vector encodes a temperature-sensitive repressor, a special *E.coli* strain was not required for regulation of expression.

The entire coding region of the MAV-RT (*Eco*RI-*Xho*I fragment, obtained by restriction digestion or PCR with suitable primer pairs), as characterized by the restriction map data of Table I, was inserted into the multiple cloning site (MCS) of pH. Briefly, the vector was restricted with *Eco*RI and *Sal*I. A 1:1 ratio of insert to vector was ligated in the presence of 1 mM ATP in ligation buffer (100 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 20 mM DTT) and T4 DNA ligase using a convention protocol. Sambrook *et al.*, *in* Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (2d Ed. 1989). The ligation mix was incubated at 16°C for 2-4 hours. The ligated mix was transformed into electro-competent *E.coli* cells in 1 mm cuvettes using a BioRad electroporator at 1.8 KeV and 200 ohms. The transformed cells were plated on LB-ampicillin plates and single colonies were picked for overnight growth and mini-prep analyses. The recombinants were then confirmed by sequence analyses. Subsequently, the 5′ noncoding regions of selected recombinants were removed by site-directed mutagenesis where appropriate. The pH vector containing the full-length RT gene was named pHSEM1 and the vector having the 5′ non-coding region deleted was called pHSEMUE33 (*i.e.*, pHRT). The RT protein was expressed and analyzed by SDS-PAGE and RT assays were performed as described in Examples 5 and 6. Other prokaryotic vectors were also successfully used (*e.g.*, pET21d and pTZ18U, which have the T7 promoter and the lacZ promoter, respectively).

Construction of Eukaryotic Recombinant Transfer Vectors

A baculoviral expression system consisting of a transfer vector, a wild-type virus AcMNPV (Autographa californica nuclear polyhedrosis virus) or a derivative of ACMNPV (*i.e.*, BacPak6 (Clontech Inc.)) was used to obtain recombinant transfer vectors containing the RT gene.

The AcMNPV genome is a double-stranded circular DNA of 134 kb. The size of the virus makes it difficult to directly manipulate the viral genome itself. Therefore, transfer vector pBacPak9 was used to generate recombinant molecules in accordance with the invention, such as pMBacRT, pBacMIBA, pBacMIKA, pBacMIBAhis and pBacMIKAhis (see below). These recombinant molecules, containing exogenous and typically foreign RT coding regions, were used to introduce the sequence into the viral genome for expression and propagation. Vector pBacPak9 has a strong polyhedron promoter which is induced in insect cells late in the replication cycle of the virus. Hence, foreign genes, including lethal genes, expressed with this late promoter are not toxic to the growing cell. The polyhedron gene is not necessary for the maintenance of the virus and was therefore replaced by the foreign gene (*i.e.*, MAV-RT *pol* gene).

The 2.81 kb *Pst*I fragment from pHSEM1, containing the full-length RT gene, was inserted into the *Pst*I site of the MCS of pBacPak9, and the recombinants were called pBpHPC3,4 (*i.e.*, pBacRT). Insertions of the gene were confirmed by miniprep analyses and sequencing. The 5′ non-coding region (see, SEQ ID NO:1) was removed by site-directed mutagenesis, as described in Sambrook *et al.*, (1989). The resulting recombinant vector was called pBpHPCM10,11,17 (*i.e.*, pMBacRT). In pMBacRT, the RT gene is flanked by viral DNA sequences of BacPak6, a derivative of AcMNPV. When pMBacRT was introduced into insect cells along with BacPak6 DNA, the plasmid recombined with the BacPak6 DNA to yield recombinant, infectious progeny virus (M1-5 and M1-6, collectively M1-5,6) containing the RT gene.

In general, when SF9 tissue culture cells are infected with recombinant virus, the viral particles entered the cells and the viral DNA is uncoated in the nucleus. Viral DNA replication occurs approximately 6-24 hours post-infection. During the late phase of the viral infection, approximately 48-72 hours after virus infection, all transcription is shut off except the genes having the polyhedron and p10 promoters, which are transcribed at very high levels. Hence, the RT gene under the control of the polyhedron promoter in the

recombinant virus was expressed at high levels late in the infection cycle. This recombinant AcMNPV was propagated in the budded form only.

Example 2

A primer walking sequencing strategy implementing Sanger's enzymatic sequencing technique was used to confirm the sequence of the MAV-RT *pol* gene. Sambrook *et al.*, (1989). The sequencing template was the insert of pHSEM1. Primers were designed to be homologous or complementary to an end of a previously determined sequence. These primers were then used to progressively extend the identification of *pol* gene sequence until the sequence of the entire coding region had been determined.

The polynucleotide sequence of the MAV-RT gene and the flanking sequences are set forth as SEQ ID NO:1. Amino acid sequences encoded thereby are set forth in SEQ ID NO:2. Of the 3,155 bp presented in SEQ ID NO:1, 2,498 bp codes for the beta fragment (nucleotides 253-2751 of SEQ ID NO: 1) of MAV-RT; the alpha fragment of MAV-RT is encoded by nucleotides 253-1990 of SEQ ID NO:1. These coding regions are expected to encode polypeptides containing amino acids 1-895 of SEQ ID NO:2 (full-length RT; see also SEQ ID NO:3), amino acids 1-833 of SEQ ID NO:2 (β-like polypeptide; see also, SEQ ID NO:5) and amino acids 1-579 of SEQ ID NO:2 (α-like polypeptide; see also, residues 1-578 of SEQ ID NO:4). The β-like polypeptide is a fragment of the native MAV-RT β polypeptide. The α-like polypeptide is larger than the native MAV-RT α polypeptide and smaller than the native MAV-RT β polypeptide, with the native α polypeptide sequence extending from the N-terminus of the α-like polypeptide. For brevity, the α-like and β-like polypeptides are referred to as the α and β polypeptides, respectively.

Example 3

As described in Example 1, plasmids pHSEM1 and pBacRT were constructed to contain 2.95 kb and 2.81 kb inserts, respectively. These fragments contained the entire reverse transcriptase gene along with 5′ and 3′ non-coding regions. The 5′ non-coding region of each construct was then removed by site-directed mutagenesis, a well-known technique in the art. In particular, the primer FSDRT (5′-TGTACTAAGGAGGTG-TTCATGACTGTTGCGCTACAT-3′; SEQ ID NO:20) was used with pHSEM1 as a template to generate pHRT (pHSEMUE33). Primer RSDBAC2 (5′-GCCAGATGT-

AGCGCAACAGTCATATTTATAGGTTTTTTTATTAC-3'; SEQ ID NO:21) was used with pBacRT as a template to generate pMBacRT (pBPHPC3M10, pBPHPC3M11, pBPHPC3M17, or, respectively, pMBac10, pMBac11 and pMBac17).

The full-length RT coding region was used as a starting material in constructing deletion derivatives that lacked the 3' end of the MAV-RT coding region to varying extents. Relative to the full-length gene (MI-5,6, see below), the 3' (C-terminal) deletion extending to the *Kpn*I site (MIKA) increased the RT expression level, as evidenced by SDS-PAGE. Relative to the full-length gene (MI-5,6), deletion of the region extending from the *Bgl*II site to the 3' terminus (MIBA) increased the RT expression level, activity and solubility, as evidenced by SDS-PAGE and activity assays (see below). Relative to the alpha fragment of MAV-RT, the beta fragment has an additional 254 amino acids at the C-terminus, which provides an integrase activity. This region of the polypeptide contributes to the insolubility of the polypeptide and reduces its recovery from cell extracts, as shown by the relative insolubility of a (+) integrase form of RT (*e.g.*, the M1KA gene product, see below) compared to a (-) integrase form (*e.g.*, the M1BA gene product). Because the integrase domain is only needed for the retroviral life cycle and not for the RNA- or DNA-dependent DNA polymerase activities, this region was deleted in M1BA (α fragment equivalent). Note that the α fragment of M1BA (amino acids 1-578 of SEQ ID NO:2) is larger than the naturally occurring α fragment of MAV-RT (amino acids 1-573 of SEQ ID NO:2). Without wishing to be bound by theory, this deletion was expected to result in an increase in the solubility, and hence recovery, of the protein.

Using the full-length RT recombinants, additional clones were constructed to express polypeptides having C-terminal deletions in order to increase the levels of expression and to stabilize the RT activity (RNA-dependent DNA polymerase activity). Convenient restriction sites such as *Bgl* II (spanning nucleotides 1,986-1,991 of SEQ ID NO:1) and *Kpn*I (spanning nucleotides 2,745-2,750 of SEQ ID NO:1) were used to eliminate the 3' end of the coding region of the RT gene (see, Table I). The 3' deletion derivatives, encoding RT polypeptide fragments having C-terminal deletions, were obtained by *Bgl*II-*Pst*I or *Kpn*I-*Pst*I restrictions of pMBacRT and pHRT, respectively (*Bgl*II and *Kpn*I sites in the MAV-RT coding region; *Pst*I site in the vector). Recombinant molecules containing the *Bgl II-Pst*I 3' terminal deletion were designated pBacMIBA and pHBRT (pH33ΔBP6) and recombinant molecules containing the *Kpn*I-*Pst*I deletion were designated pBacMIKA and pHKRT (pH33ΔKP5). The deletion derivatives pBacMIBA and

pBacMIKA had approximately 1.2 and 0.4 kb deletions from the 3′ end of the full-length gene (see, SEQ ID NO:1), respectively. The fragment bounded at its 3′ end by the *Bgl*II site (SEQ ID NO:6) was used to express an alpha fragment equivalent of RT and the fragment bounded by the *Kpn*I site (SEQ ID NO:8) was used to express the beta fragment equivalent of RT (the β fragment equivalent of M1KA contained amino acids 1-832 of SEQ ID NO:2; native MAV-RT β contains amino acids 1-858 of SEQ ID NO:2).

Miniprep and sequencing analyses were done to confirm the identities of the recombinant clones described above. Recombinant viruses obtained from co-transfection with virus BacPak6 and transfer vector pBacMIBA or pBacMIKA were called M1BA and M1KA, respectively.

Recombinants encoding 3′ terminal amino acid tags

Without wishing to be bound by theory, the constructs that deleted the integrase domain of RT, such as M1BA and pBacMIBA, were not expected to retain the DNA binding, structure stabilizing, and polymerization functions attributable to the integrase domain. To re-introduce these functions, without the deleterious impact on solubility and host cell viability associated with the native integrase domain, codons specifying amino acids (His) were added to the 3′ end of the modified RT coding regions. The basic nature of the added amino acids may have been responsible for increased binding to the negatively charged nucleic acids, enhancing the stability of the polypeptides. The increased binding may, in turn, have been responsible for the increase in activity found with the his-tagged RTs, relative to their untagged counterparts. In addition, the his tags may have contributed to the tendency of the his-tagged RTs of the invention to form polymers, perhaps through his mediated chelation of metal ions such as Ni^{++}. A his-tagged RT (MIBAhis) was found in homo-polymeric form (molecular weight greater than 200 kDa), as determined using non-denaturing PAGE and molecular sieve chromatography with Superose 12HR10/30 (separation range of 1-300 kDa; Pharmacia-Upjohn). Thus, the invention contemplates RT polypeptides lacking an effective integrase domain, but having the capacity to bind DNA and/or polymerize. These additional functionalities may be provided by adding, preferably at the C-terminus of the modified RT, such structures as known DNA binding domains, zinc-finger or zinc-finger-like domains, polymerization domains, acidic amino acids, basic amino acids, or one or more cysteines. Such modified RTs may be ultimately derived from avian or non-avian sources.

His-tag additions to the C-termini of the RT polypeptides were achieved by recombinant expression of coding regions fusing RT coding regions to His codons. In particular, the fusions were constructed by adding oligonucleotides containing 6 histidine codons at the 3′ end of the RT gene using ligase, as in the case of the construction of pBacMIKAhis, or by PCR amplification with oligonucleotides that specified 6 histidine codons, as in the case of the construction of pBacMIBAhis.

The construction of pBacMIKAhis was accomplished with oligonucleotides FNhis (SEQ ID NO:33) and RNhis (SEQ ID NO:34), each of which contained internal histidine codons and compatible *NotI* restriction sites at each end. Following their conventional syntheses, the oligonucleotides were annealed and ligated to the 3′ terminus of RT in pBacMIKA cut with *Not*I. For the construction of pBacMIBAhis or pBacMIKAhis using PCR, primers FRT (SEQ ID NO:22) and either M1BARSDhis (SEQ ID NO:23) or M1KARSDhis (SEQ ID NO:24) were used with pHSEM1 as the template. Blunt-ended and phosphorylated PCR products containing the 3′ deletions and histidine tag-encoding regions were inserted into the *Sma*I site in the MCS of pBacPak9. The his-tag derivatives of the transfer vectors were called pBacMIBAhis and pBacMIKAhis and the viruses obtained by co-transfection of Sf9 cells with the aforementioned transfer vectors and BacPak6 were called M1BAhis ((-) integrase) and M1KAhis ((+) integrase), respectively. Introduction of the His codons led to increased activity of the encoded polypeptides in eukaryotes, as measured by SDS-polyacrylamide gel electrophoretic analyses and RT assays (see below). As shown below, the his-tag additions increased the stability (perhaps by providing a DNA binding site), activity, polymerization capabilities and ease of purification of RTs such as M1BAhis.

The 5′ end of the MAV *pol* gene was also modified. Beyond deletion of the 5′ non-coding sequence of *pol* (see the description of pHRT and pMBac10 above), the widely recognized Met initiation codon ("ATG") was introduced immediately upstream of the natural start codon (the Thr codon "ACT" at nucleotides 253-255 of SEQ ID NO:1) of the MAV *pol* gene.

In general, the above-described cloning strategy reflected efforts to eliminate the integrase domain of avian RT and thereby avoid the insolubility and lethality problems associated with that protein domain. Deletion of 192 bp from the 3′ terminus of the full-length MAV-RT gene (SEQ ID NO:1) by terminating the coding region at the *Kpn*I site (Table I) produced the "MIKA" clone series. These clones coded for a β polypeptide that

is smaller than the naturally occurring β polypeptide. These clones exhibited enhanced RT expression and the expressed polypeptides exhibited enhanced activity levels (compare below, the expression of M1-5,6 [full-length] to M1KA [β polypeptide]). Larger deletions extending from the 3′ end of the full-length MAV-RT gene were constructed using a convenient *Bgl*II site to generate the M1BA clone series. These clones encoded an α subunit of RT that was larger than the naturally occurring α polypeptide. The M1BA clones exhibited increased expression and activity, in comparison to the expression and activity of full-length MAV-RT; moreover, M1BA was more soluble than naturally occurring MAV-RT.

The invention also contemplates polynucleotides and polypeptides resulting from a recognition that some advantageous properties of the integrase, *e.g.*, DNA binding and polymerization, could be re-introduced into avian RTs without re-introducing the deleterious (*i.e.*, insolubility and lethality) characteristics of the avian RT integrase domain. One approach is to attach RT integrase domains or non-RT integrase domains known in the art to the (-) integrase polypeptides or attach the coding regions of these domains to the polynucleotides encoding these (-) integrase polypeptides. Another approach is to add amino acid tags to the (-) integrase RT polypeptides (or corresponding codons to (-) integrase polynucleotides) as disclosed herein. A preferred tag is a basic amino acid tag such as a His tag. As disclosed below, a His tag was attached at the C-terminus of an α polypeptide equivalent (MIBAhis). This clone exhibited relatively high levels of expression, activity and solubility. Thus, the invention provides avian RTs improved in terms of expression and activity levels, and in terms of solubility and ease of purification, while retaining the processivity and thermostability characteristic of avian RTs.

Accordingly, the invention contemplates the construction of analogous polynucleotides and recombinant molecules encoding RT polypeptides of unnatural length from other sources, such as MMLV, HIV, RSV, ASLV, ATV, and others. Further, the invention extends to polynucleotides encoding these RTs of modified length, or full length RTs, provided that the polynucleotides additionally encode polymerizing or nucleic acid binding domains, and preferably both domains, at their 3′ termini. Examples of polynucleotides encoding a non-avian RT of unnatural length are polynucleotides encoding an RT portion or fragment having the amino acid sequence set forth at any one of the following: positions 1-765 of SEQ ID NO:39 (an HIV-2 RT sequence), positions 1-800

of SEQ ID NO:41 (an MMLV-RT sequence), and positions 1-625 of SEQ ID NO:43 (an HIV-1 RT sequence). These polynucleotide sequences have some correspondence to the sequence of the polynucleotide encoding the MAV-derived M1BA polypeptide and are expected to function in a manner analogous to polynucleotides encoding M1BA. Of course, a polynucleotide encoding the full-length β polypeptide of HIV-2 (SEQ ID NO:38), or encoding equivalent polypeptides from MMLV or HIV-1 (SEQ ID NO:40 or SEQ ID NO:42, respectively), along with a 3′ terminal sequence encoding a polymerizing and or nucleic acid binding domain, are also contemplated by the invention.

With respect to polypeptides, the invention comprehends the polypeptides encoded by the above-described polynucleotides, as well as polypeptides that have a C-terminal polymerizing and/or nucleic acid binding domain that has been added by means other than expression. For example, an RT polypeptide having a Cys residue or a His residue attached at the C-terminus by chemical condensation falls within the scope of the present invention. In addition, effective elimination of an integrase domain, such as is found in avian RTs, may be effected by altering a suitable coding region by inserting, deleting, or substituting (transitions and/or transversions), one or more nucleotides. Thus, the invention contemplates RT polypeptides that are the same length as naturally occurring RT polypeptides. These RT polypeptides may have the same amino acid sequence as naturally occurring RTs, provided that the RTs of the invention have a polymerizing and/or nucleic acid binding domain at their C-termini. Alternatively, RTs of the same length as natural RTs may have sequences that differ from the natural RTs, thereby effectively eliminating integrase activity. The RTs of the invention may also be shorter, or longer, than naturally occurring RT polypeptides. The shorter RT polypeptides of the invention eliminate some, or all, of the C-terminal sequence of a naturally occurring RT which, in the case of avian RTs, contains the integrase domain. RTs of the invention that are longer than naturally occurring RT polypeptides contain the sequence of that naturally occurring RT and, in addition, sequence of an adjacent peptide region. Additionally, these polypeptides of unnatural length may have a polymerizing and/or nucleic acid binding domain added at their C-termini.

Example 4

The RT constructs described in Example 3 were transformed into prokaryotic and eukaryotic host cells and the expression of RT polypeptides was analyzed. A prokaryotic host cell, *Escherichia coli* DH5αF′, was transformed with pHRT, pHBRT or pHKRT, using a technique standard in the art. Cells subjected to the transformation protocol were plated on LB plates (10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 ml of 1N NaOH, 1.5 g agar, ddH_2O in a total volume of 1 liter) containing 50 μg/ml ampicillin for selection of transformed host cells. Single colonies were picked, expanded in small culture (*i.e.*, 5 ml), episomal DNAs were rapidly isolated from an aliquot of cells, and the purified DNAs were analyzed for the presence of a recombinant molecule of the expected size. Dideoxynucleotide-based sequencing of these DNAs confirmed that the first ATG (*i.e.*, the initiation codon) was in-frame with the remainder of the RT coding region.

Another aliquot of those small cultures containing cells transformed with pHRT, pHBRT, or pHKRT was used to inoculate flasks containing 10 ml of LB-ampicillin and grown at 30°C until an OD_{600} of 0.6 was reached. Flasks containing these cells were then quickly shifted to 42°C to de-repress the λP_R promoter and express the recombinant protein. After an hour at 42°C, cells were pelleted and analyzed for expression of protein by SDS-PAGE, Western blot analyses, and RT activity assays, as described below.

In general, about 10% of the expressed protein was recovered in soluble form and 90% of the expressed protein was found in inclusion bodies, as revealed by pelleting lysed cells at 12,000 x *g* for 5-15 minutes. RT activity was also found when expressing both the full-length and the deletion derivatives of the MAV *pol* coding region from other recombinant vectors, such as pTZ18U and pET21d, that contained similar insert fragments encoding full-length or C-terminally deleted MAV-RT.

A eukaryotic host cell suitable for use in practicing the invention is the Sf9 insect cell. Several polynucleotides were separately introduced into Sf9 cells using the Baculoviral expression system. O'Reilly *et al.*, *in* Baculovirus Expression Vectors: A Laboratory Manual, Oxford University Press (1994). The polynucleotides (*i.e.*, pMBacRT, pBacMIBA, pBacMIBAhis, pBacMIKA, and pBacMIKAhis) were purified by the standard alkaline lysis method, as described in Sambrook *et al.*, (1989). The DNA was then centrifuged through a CHROMA SPIN+TE-400 column (Clontech Laboratories, Inc.,

Palo Alto, CA.) at 500 x *g* for 7 minutes in a swinging bucket rotor. (HN-SII centrifuge from IEC, Inc.) This purified DNA was then used to transform eukaryotic cells.

Sf9 insect host cells were prepared for transformation using an established procedure. The Sf9 cells from an exponentially growing cell culture were initially counted using a hemocytometer and diluted to $5x10^6$ cells/ml of TNM-FH Insect Cell Medium (Product No. T-1032; Sigma Chemical Co., St. Louis, MO.) with 10% fetal bovine serum (FBS) and antibiotics (50 units/ml nystatin, 50 units/ml penicillin, and 50 μg/ml of streptomycin). Subsequently, 1.5 ml of this culture was added to each well of several 12-well tissue-culture plates. The cells were allowed to attach to the plate for a period of 1 hour. The medium covering the cells was then removed and 2 ml of TNM-FH medium without serum was added. The serum-free medium was swirled over the cells and again the medium was removed. This process was repeated one more time to remove all traces of fetal bovine serum (*i.e.*, FBS) and antibiotics. The cells were then incubated in TNM-FH medium for 30 minutes while the transfection mixture was prepared.

The 50 μl transfection mixture contained 500 ng of DNA, 500 ng of *Bsu*36I-digested BacPak6 viral DNA, and ddH_2O. This mixture was gently mixed with 50 μl of transfection reagent (Clontech, Inc.) and incubated at room temperature for 15 minutes to allow the transfection reagent to form a complex with the DNA, as recommended by the supplier of the transfection reagent.

Medium covering the Sf9 cells was removed and 300-500 μl of TNM-FH medium was added to each well. To this medium, the transfection reagent-DNA mixture was added drop-wise while gently swirling the dish. The cells were then incubated at 27°C for 5 hours before adding 2 ml of TNM-FH medium containing 10% FBS and the antibiotics identified above. DNA-cell contact was continued at 27°C for 60-72 hours. Medium from these plates was then collected and used as primary virus stocks.

Primary virus stocks were subsequently subjected to plaque purifications by standard methods, as described in King *et al.*, *in* The Baculovirus Expression System: A Laboratory Guide (eds. Chapman and Hall, N.Y. 1992), to produce clonal stocks. The clonal stocks were amplified using a 1:1 virus to insect cell ratio to produce large quantities of recombinant viruses.

The viruses from the clonal stocks were used to infect insect cells and ultimately analyze RT expression in a eukaryotic environment. Based on the titer obtained from the plaque assays, an infection was set up using a ratio of 5 viruses per Sf9 cell. After 60

hours, the medium and cells were collected. The cells were pelleted, resuspended in cell lysis buffer (10 mM Tris HCl, pH 8.0, 50 mM NaCl, 5% glycerol, 0.5% Triton X-100, and protease inhibitors (50 μg/ml Benzamidine HCl, 0.1 mM 4-(2-aminoethyl)-benzene-sulfonylfluoride, and 1 μg/ml pepstatin A)) and lysed by sonication. These samples were subsequently subjected to SDS-PAGE, Western blot analyses, and RT activity assays.

For large-scale expression studies, Sf9 cells were initially grown in T25 tissue culture flasks under the conditions described above. Sf9 cells adhering to the T25 tissue culture flasks were gently dislodged and adapted to suspension cultures as described by King *et al.*, 1992. These suspension cultures were expanded in spinner flasks to a volume of 1-3 liters. When the insect cells reached a density of $1x10^6$ cells per ml, they were infected with a concentrated stock of recombinant viruses at a ratio of 5:1 viruses per insect cell. A variation of a standard protocol was used to infect these cells. A large volume of amplified viral stock (M1BA, M1KA, M1BAhis, and M1KAhis, or M1-5,6) was concentrated using one-half volume of 40% PEG 8000 and one-sixth volume of 5 M NaCl. Precipitated viruses were collected at 12,000 x *g* for 30 minutes using a Sorvall RC5C centrifuge (Dupont, Newtown, CT). The pelleted viruses were resuspended in 1x PBS (10 mM $K•PO_4$, pH 7.5, and 150 mM NaCl) at 1/20 of the culture volume and stored at -20°C. Before infection, the viruses were filtered through a 0.2 μ filter.

After a 48 hour period of infection, 1 ml aliquots of infected cells were collected for RT assays to monitor RT expression levels. Cells were harvested at the peak of RT expression (generally around 60 hours post-infection), as determined from previous trials. Cells were pelleted at 5,000 x *g* for 30 minutes and stored at -80°C.

Polypeptides expressed in insect cells were also characterized by SDS-PAGE and Western blot analyses. Results of a Western blot assay using a mixture of anti-RT monoclonal antibodies 1D8, 2E10, 6F1, 4C4, 9H10 and 9C2 are shown in Fig. 1 (lane 1: prestained molecular weight markers of 123 kDa, 90 kDa, 64 kDa, 50 kDa, and 38 kDa; lane 2: native AMV-RT(nRT) (lane 2), lane 3: M1BAhis, and lane 4: M1KAhis. Further analysis of the antigenic properties of M1BAhis and native RT revealed that monoclonal antibody 6F1 recognized native RT but failed to recognize the M1BAhis polypeptide. Thus, at least one epitope found on native RT is not found on M1BAhis, indicative of structural differences between the proteins.

The results further indicate that both M1BA and M1KA expressed ten-fold more RT than M1-5,6, which encodes full-length RT. When cell pellets were assayed for RNA-

dependent DNA polymerase activity, M1BA was expressed at 10,000 units per liter of insect cell culture, whereas M1KA and M1-5, 6 were each expressed at 1,000 units per liter of insect cell culture. Though M1KA expressed as well as M1BA when analyzed on Western blots, active M1KA recovered from the cell pellet was ten-fold less than M1BA. Most of the expressed M1KA remained insoluble in the pellet. Although the corresponding his-tagged proteins (M1BAhis and M1KAhis) were expressed at levels similar to their M1BA and M1KA counterparts as revealed by Western blotting, the activities of the his-tagged proteins were higher. M1KAhis was expressed at 2,000 units per liter of insect cell culture and M1BAhis was expressed at 200,000-400,000 units per liter of insect cell culture.

The Baculoviral system is preferred for expression of RT and fragments thereof. A relative comparison of RT expression in prokaryotic and eukaryotic cells, as measured by reverse transcriptase assays of purified recombinant and crude protein, revealed that His-tagged RT polypeptides from eukaryotic insect cells were most active and stable, while untagged polypeptides expressed in prokaryotic cells were less active and stable.

Recombinantly expressed polypeptides of the invention were purified using conventional protocols, with metal-affinity chromatography included for the isolation of His-tagged polypeptides. Host cells containing recombinant molecules (*i.e.* M1-5,6, M1BA, M1KA, M1BAhis and M1KAhis) encoding an RT or fragment thereof were centrifuged and the cell pellet was solubilized in 20 ml cell lysis buffer (20 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.5% Triton X, and 5% glycerol) per gram of cell pellet. The resuspended cells were sonicated with five 30-second bursts at 50% power on ice with 30 seconds of cooling between each round of sonication. Sonicated cells were then stirred at a low speed on a magnetic stirrer at 4°C for one hour to complete cell lysis. The lysed samples were centrifuged at 12,000 x *g* for 30 minutes. The pellet was discarded and the supernatant was subjected to column chromatography.

RTs lacking his tags were purified according to conventional protocols, which included removal of cell debris by centrifugation and subjection of supernatants to chromatographic purification procedures known in the art. The soluble extract containing his-tagged RTs were mixed with a commercially available Ni^{++} affinity column (Ni-NTA resin from Qiagen, Inc., Chatsworth, CA), thereby using the his tags for their known purpose of facilitating purification via metal affinity chromatography. The extract and affinity resin were gently rocked on ice for 1 hour in a 50 ml plastic test tube. The resin

was then packed in a column and washed with two column volumes of wash buffer (20 mM Tris HCl, pH 8.0, 250 mM NaCl, 0.5% Triton X-100, and 5% glycerol) and two column volumes of buffer A (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.5% Triton X-100, 5% glycerol and 50 mM imidazole). (Of course, the extract could have been applied to a pre-formed affinity column and purified using conventional column chromatography, as would be understood in the art.) The protein bound to the column was eluted by setting up a linear gradient from buffer A to buffer B (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.5% Triton x 100, 5% glycerol and 250 mM imidazole).

Fractions from the nickel affinity column that had RT activity were analyzed by SDS-PAGE to determine the purity of the protein, as shown in Fig. 2. Fig. 2 presents an electrophoretogram of an 8% SDS-PAGE gel stained with Coomassie Blue. The lanes of the gel shown in Fig. 2 contain molecular weight markers of 94 kDa, 64 kDa, 43 kDa, 30 kDa and 20 kDa (lane 1) and aliquots of fractions obtained from the nickel affinity column (lanes 2 to 4). The fractions that were greater than 95% pure were pooled and dialyzed against storage buffer (200 mM KPi, pH 7.2, 5 mM DTT, 0.2% Triton X-100 and 50% glycerol). Additionally, conventional purification steps may be incorporated into the protocol to achieve greater purity, as would be understood in the art.

Protein concentrations were determined using the Bradford protein assay (BioRad Laboratories, Inc., Hercules, CA). Generally, the specific activity of rRT (MIBAhis) was calculated to be approximately 30,000-100,000 units/mg, which is similar to the specific activity of nRT (30-100,000 units/mg).

Example 5

The purified rRT prepared from cultures expressing M1BAhis at 400,000 units/liter of culture, a level well beyond a commercially feasible production limit, was found to be greater than 95% pure as judged by electrophoretic fractionation using 10% SDS-PAGE. The apparent molecular weight of the monomer is 60 kDa, which compares well with the calculated molecular weight of approximately 59.5 kDa. The recombinant protein was analyzed on a 12.5% polyacrylamide non-denaturing gel for the presence of monomers and polymers (*e.g.*, dimers) using the Pharmacia Phast System. The protein sample was prepared in either of two ways. One aliquot was completely denatured by heating at 100°C for three minutes in treatment buffer (0.125 mM Tris-HCl, pH 6.5, 4% SDS, 20% glycerol, 10% β-mercaptoethanol). Another aliquot was partially denatured at 70°C in

treatment buffer without 2-mercaptoethanol. Under completely denaturing conditions, rRT was observed to migrate at approximately 66 kDa (BSA marker) and the partially denatured samples had additional bands ranging from 60-200 kDa, indicating that rRT formed polymers. Protein size determinations were confirmed using molecular sieve chromatography with Superose 12HR10/30 (separation range of 1-300 kDa), as described above, which revealed that the majority of the rRT eluted between beta amylase (approximately 200 kDa) and apoferritin (443 kDa).Thus, the rRT was predominantly in a polymeric form. Without wishing to be bound by theory, the addition of C-terminal histidine residues may have provided a polymerization capacity, perhaps by complex formation via metal (*e.g.*, nickel) chelation, to substitute for the loss of that capacity attributable to the integrase domain, which had been deleted. Thus, the invention contemplates RT polypeptides having C-terminal attachments in the form of compounds capable of promoting polymer formation. Suitable compounds would include, but are not limited to, a plurality of basic or acidic amino acids, as well as Cys residues capable of disulfide bond formation.

Expressed rRT was also characterized immunochemically. Monoclonal antibodies against AMV reverse transcriptase were prepared using techniques well known in the art. See Harlow *et al.*, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1988). Briefly, spleen cells from a mouse that had been immunized with RT were fused with mouse myeloma cells to make hybridomas. These hybridomas were allowed to grow into colonies in 96-well plates; supernatants from these wells were then tested to find hybridomas that appeared to make anti-RT antibodies. Further testing confirmed these results.

To prepare spleen cells for hybridoma production, a BALB/C mouse (female, ten weeks old, obtained from Harlan Sprague Dawley, Madison, WI) was immunized by several intraperitoneal injections with AMV-RT (Molecular Biology Resources, Inc.) using a conventional immunization schedule. To prepare RT for injection, the storage buffer was removed from purified RT by diluting the enzyme in phosphate-buffered saline (PBS) and reconcentrating it using a Centricon 30 concentrator (Amicon Corp.). The concentrated RT was then diluted again in PBS and emulsified with an equal volume of an adjuvant. For the initial injection, the adjuvant was complete Freund's adjuvant (Sigma Chemical Co.); for the booster injections, the Ribi Adjuvant System (Ribi Immunochem Research, Inc. Hamilton, VT) was used. The dose of RT was approximately 20 micrograms per

injection. The injections were made over a period of eight months, with successive intervals of five weeks, four weeks, three weeks, eleven weeks, eight weeks, and three weeks. The fusion was performed five days after the final boost.

For the fusion experiment, the mouse was sacrificed and spleen cells were isolated and fused with myeloma cells (P3X63-AG8.653, ATCC CRL 1580), using procedures well known in the art. See Harlow *et al.* In particular, the cells were fused in 50% polyethylene glycol, resuspended in a selection medium (*i.e.*, HAT medium), and distributed into the wells of fourteen 96-well plates. After three weeks of growth, approximately 350 wells contained hybridoma colonies.

Hybridomas making anti-RT antibodies were identified by ELISA. For this procedure, the wells of 96-well polystyrene ELISA plates were first coated with purified RT (2 micrograms RT/ml in 100 mM Tris-HCl, pH 8.5, 0.05% NaN_3; overnight incubation at room temperature), then washed with TBST (Tris-buffered saline, pH 8.5, 0.05% Triton X-100) to remove excess RT. For the assay itself, the wells were filled with 95 microliters of TBST plus 5 microliters of hybridoma culture supernatant. The plates were incubated at room temperature for two hours, then washed with TBST to remove unbound immunoglobulin. To detect wells with anti-RT antibodies, peroxidase-conjugated goat anti-mouse IgG (heavy-chain specific; Jackson ImmunoResearch, West Grove, PA) was diluted 5,000-fold into TBST and added to the wells of the ELISA plates. After the wells had been incubated for one hour at room temperature, the unbound peroxidase conjugate was removed by thorough washing of the plates with TBST. Wells positive for RT were visualized colorimetrically following addition of the substrate 3-methyl-2-benzothiazolinone hydrazone/3-dimethylaminobenzoic acid/hydrogen peroxide to detect immobilized HRP. Hybridomas from positive wells were repeatedly cloned by limiting dilution until all wells with growth were ELISA-positive.

Supernatants from wells that tested positive by ELISA were further screened by immunoprecipitation of RT using techniques well known in the art. See Harlow *et al.* The immunoprecipitation assay relies on the presence of protein A (which binds IgG) on the surface of *Staphylococcus aureus* cells (SAC, Sigma Chemical Co.). Since protein A does not bind strongly to mouse IgG, a pellet of centrifuged SAC cells was first treated with rabbit anti-mouse IgG antibodies. The pellet from 10 microliters of a 10% suspension of these cells was then incubated with 50 microliters of hybridoma culture supernatant for 2 hours at room temperature. The resultant SAC cells were centrifuged, washed, and

resuspended in diluted RT. The RT cell suspensions were incubated for 3 hours at 4°C and centrifuged. The resultant supernatants were removed and tested for depletion of RT activity using a standard radiochemical assay.

Six hybridoma lines tested positive in both the ELISA and immunoprecipitation assays. These lines were designated 1D8, 2E10, 4C4, 6F1, 9C2 and 9H10. All six monoclonal antibodies had gamma-1, kappa isotypes.

The form of active rRT (*i.e.*, monomer or polymer) was confirmed using ELISA in a sandwich format with anti-RT monoclonal antibodies. Initially, monoclonal antibody was immobilized in DNA bind plates, Costar Corp., Cambridge MA. The plate was then blocked with BSA to prevent non-specific binding. The wells were then incubated with purified rRT (*i.e.*, M1BAhis). Excess or unbound protein was removed by washing with phosphate-buffered saline. The wells were then incubated with the same monoclonal antibody linked to biotin for detection. If the rRT existed as a monomer, the biotin-linked monoclonal antibody should not bind to it. However, the biotin-linked monoclonal antibody did bind to the rRT, indicating that the rRT had formed a polymer.

To determine the purity of the samples containing reverse transcriptase, recombinant protein expressed from each of a variety of clones (*e.g.*, M1BAhis) and found in either the solubilized cell pellets or protein fractions from the different chromatographic columns used in purification were subjected to SDS-PAGE. Samples were electrophoresed on 8% polyacrylamide gels containing 6% stacking gels, followed by Coomassie Blue R-250 staining using standard protocols (Sambrook *et al.*, 1989). The recombinant protein was found to be greater than 95% pure.

Using the Pharmacia Phast System, the recombinant (MIBAhis) and native reverse transcriptase, as well as appropriate standards supplied with the system (i.e., IEF 3-9), were subjected to isoelectric focusing electrophoresis. (Pharmacia-Upjohn, Piscataway, NJ.) The experimentally determined pI values of the rRT and rRT were 6.0. The theoretical pI of rRT, calculated from its amino acid sequence, was 6.8.

For analyses of total expression, host cells containing one of several recombinant DNAs (*i.e.*, pMBacRT, pBacMIBA, pBacMIKA, pBacMIBAhis, and pBacMIKAhis) were induced to express recombinant protein. The induced cells were pelleted at 12,000 x *g* for 5 minutes. The cell pellet was then resuspended in SDS sample buffer (Sambrook *et al.*, 1989) or cell resuspension buffer (20 mM Tris HCl, pH 8.0, 250 mM NaCl, 0.5% Triton X-100, and 5% glycerol) to assess the solubility of the protein. Resuspended cells were

pulse-sonicated three times at a setting of 3 (Virsonic 100 from Virtis Company, Inc., Gardiner, NY) for 20 seconds each (500 mM Tris HCl, pH 6.5, 14% SDS, 30% glycerol, 9.3% DTT, and 0.012% bromophenol blue). Small aliquots of the samples in SDS sample buffer were loaded on duplicate gels and electrophoresed. One of the duplicate gels was stained with Coomassie Blue and the other gel was used to transfer protein to a 0.2 μ nitrocellulose membrane using a Bio-Rad transfer apparatus for Western blot analysis. Bio Rad Laboratories, Inc., Hercules, CA. Detection of expressed protein in fractionated crude lysates was possible using specific, monoclonal anti-RT antibodies (a mixture of monoclonal antibodies 4C4, 1D8, 2E10, 6F1, 9H10, and 9C2; Molecular Biology Resources, Inc., Milwaukee, WI) to detect the recombinant protein.

In practice, the nitrocellulose membrane containing the transferred protein was contacted with a blocking buffer (5% casein hydrolysate, 150 mM NaCl, 10 mM Tris HCl, pH 8.0) for 30 minutes followed by incubation with a 1:1000 dilution of anti-RT monoclonal antibody in blocking buffer. After overnight incubation, blots were rinsed in wash buffer (10 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.5% Tween 20) and incubated with a 1:5000 dilution of alkaline phosphatase-conjugated goat anti-mouse antibody in blocking buffer for 1 hour. Subsequently, the blots were rinsed 3x with wash buffer and 1x with AP buffer (100 mM Tris HCl, pH 9.5, 5 mM $MgCl_2$ and 100 mM NaCl). RT was indirectly detected by performing a colorimetric phosphatase assay using a standard substrate mixture of NBT (nitroblue tetrazolium; 75 mg/ml in dimethylformamide) and BCIP (5-bromo-4-chloro-3-indolyl phosphate, 50 mg/ml in dimethylformamide), which forms a blue precipitate when dephosphorylated by any immunologically immobilized phosphatase. The anti-RT antibody recognized two bands, one at approximately 61 kDa and one at 92 kDa, in the lane containing native RT. In the lane containing recombinant, His-tagged RT expressed from M1BAhis (alpha fragment equivalent), a single band at approximately 60 kDa was found; in the lane containing recombinant, His-tagged RT expressed from M1KAhis (beta fragment equivalent) a single 91 kDa band was found.

Assays were also performed to determine the intrinsic/extrinsic exonuclease, endonuclease, (*i.e.*, nicking) DNase, and RNase activities of the rRT. An assay for 3'--> 5' exonuclease activity was performed using radiolabeled *Taq*I fragments of lambda DNA as a substrate. The 3' ends of *Taq*I-digested lambda DNA fragments (265 μg) were labeled with 60 μCi [^{3}H]-dCTP (57.4 μCi/mmole) and 60 μCi [^{3}H]-dGTP (8.9 μCi) using 40 units of exo$^-$ Klenow fragment of DNA polymerase in a standard labeling reaction.

Sambrook *et al.*, (1989). The 3'-->5' exonuclease assay was performed in a final volume of 10 μl containing 50 mM Tris HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM DTT, 0.015 μg of labeled *Taq*I fragments of λ DNA, and either 2.5 or 10 units of RT enzyme. One unit of RT enzyme is the amount of enzyme required to incorporate 1 nmol of dTTP into an acid-insoluble form in 10 minutes at 37°C under the stated assay conditions (see, Example 6). Each sample was incubated at 37°C for 1 hour. The reaction was terminated by the addition of 50 μl of yeast tRNA and 200 μl of 10% trichloroacetic acid. After incubation for 10 minutes on ice, the samples were centrifuged for 7 minutes in a microcentrifuge. The supernatant (200 μl), which contained the released label, was removed and added to 6 ml of scintillation fluid and counted in a scintillation counter. The results showed that the rRT released 0.13% of the label, an acceptably low level of 3'-->5'exonuclease activity.

The rRT was also subjected to a 5'-->3' exonuclease assay, using radiolabeled *Hae*III fragments of λDNA. The λ fragments were 5' end-labeled using 60 μCi [γ-^{33}P] dATP (2,000 Ci/mmol) and 40 units of T4 polynucleotide kinase in a conventional procedure. Sambrook *et al.*, (1989). Except for the use of 5' end-labeled *Hae*III fragments as substrate, this assay was performed in accordance with the description of the 3'-->5' exonuclease assay above. The purified rRT released - 0.36% of the label into an acid-soluble form, an acceptably low level of 5'-->3' exonuclease activity.

Double-stranded and single-stranded DNase assays were also performed using the protocol for the 3'--> 5' exonuclease assay, again with the exception of the type of labeled substrate being used. For each of the DNase assays, intact lambda DNA (0.5 μg) was labeled with 30 μCi [α-^{33}P] dATP (2,000 Ci/mmol) using the random primer extension technique understood in the art. Each assay used 0.015 μg of labeled λ DNA. For single-stranded DNase assays, the labeled λDNA fragments were further subjected to heat denaturation (3 minutes at 100°C followed immediately by chilling on ice) to prepare the substrate. Again with the exception of the type of substrate employed, each of the DNase assays were conducted as described above in the context of the 3'-->5' exonuclease assay. The rRT released 0.5% of the label in the double-stranded DNase assay; 0.02% of the label was released in the single-stranded DNase assay. Both results indicate acceptably low levels of DNase activities. The purified rRT was also subjected to an endonuclease, or nicking, assay by examining the extent to which the rRT converted a supercoiled substrate in the form of pBR322 to a relaxed form, as visualized by agarose gel electrophoretic

fractionation. The assay for endonuclease activity was performed in a final volume of 10 μl containing 50 mM Tris HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM β-mercaptoethanol, 0.5 μg pBR322, and 2.5, 5 or 10 units of enzyme. Each sample was incubated at 37° C for 1 hour. Two microliters of 0.25% bromophenol blue, 1 mM EDTA and 40% sucrose were added to stop the reaction. After a brief centrifugation, 6 μl of the sample were electrophoresed on a 1.0% agarose gel in 1X TBE. Sambrook *et al.*, (1989). The results showed that less than 10% of the supercoiled substrate was converted to a relaxed form, an acceptably low level of nicking activity.

The rRT was also characterized in terms of its RNase activity. In particular, this assay was designed to measure general RNase activity and, specifically, not an RNase H activity. Substrate was prepared using run-off transcription from a T7 promoter in the presence of [α-^{33}P] UTP. In particular, the plasmid pPV2 (a pTZ-based vector containing a ColE1 ori; an ampicillin selectable marker; T7, T3 and lac promoters; and a 695 bp insert from plum pox virus) was linearized with *Pvu*II. The run-off transcription reaction was performed with 1 μg of linearized pPV2, 30 μCi of [α-^{33}P] dATP (2,000 Ci/mmol), and 10 units of T7 RNA polymerase using a conventional procedure. The RNase assay was then performed in the presence of single-stranded RNA substrate (0.15 μg) and rRT (2.5, 5 or 10 units). Released label was again recovered as acid-soluble material using the TCA precipitation procedure described above. Scintillation counting showed that 1% of the radiolabel was released, indicating an acceptably low level of RNase activity.

Example 6

The RNA-dependent DNA polymerase activities of native RT and recombinant RT (purified expression product of M1BAhis) were compared. One unit of enzyme was compared in RT assays with either poly rA:dT_{12-18} (20:1) or mRNA as substrate. Product quantity was determined by either glass filter precipitation or binding to DE52 filters; product quality was monitored by autoradiography of a 1.2% TBE agarose gel containing fractionated reaction products.

The reverse transcriptase activities of the native and recombinant proteins were compared using a modification of a procedure described by Meyers *et al.*, Biochemistry 30:7661-7666 (1991). The reaction mixture contained 1x reaction buffer (50 mM Tris-HCl, pH 8.3, 40 mM KCl, 10 mM $MgCl_2$), 1 mM DTT, 0.4 mM poly rA:dT_{18}, 0.5 μCi [α-^{32}P] TTP (3,000 Ci/mmol), 0.5 mM dTTP, 1 unit of enzyme (one unit of RT enzyme

is the amount of enzyme required to incorporate 1 nmol of dTTP into an acid-insoluble form in 10 minutes at 37° C under the stated assay conditions), and ddH_2O to 50 µl total volume. Reaction mixtures without enzyme were pre-incubated at 37° C for 1 minute prior to the addition of enzyme. Reactions were then incubated at 37°C for 20 minutes, and terminated by adding 2 µl of 0.5 M EDTA followed by applying 40 µl of each reaction mixture to separate DE52 filter membranes. The filters were washed three times with 5% Na_2HPO_4 for 5 minutes each, then rinsed with ddH_2O followed by 95% ethanol. The filters were air dried, placed in scintillation fluid, and immobilized radioactivity was quantitated.

A variation of the filter assay was used to compare the quantity and quality of reaction products. Messenger RNA, 891 bp control and 7.5 kb mRNA, were obtained from GIBCO BRL, Gaithersburg, MO. The following substitutions in the assay described above were made: 1 µg of mRNA primed with 0.5 mM oligo dT_{12-18} primer, instead of poly rA:dT_{18}; and mixed dNTPs (0.5 µM each of dGTP, dATP, TTP and dCTP, and 0.02 µCi [α-^{32}P]-dATP (6,000 Ci/mmol)), instead of [α-^{32}P] dTTP. Reactions were initiated by adding 5 units of RT to the reaction mixture. After 1 hour of incubation at 37°C, a 20 µl sample was removed and mixed with 5 µl of stop solution (95% de-ionized formamide, 10 mM EDTA, 0.05% xylene cyanol FF, and 0.05% bromophenol blue) and loaded onto a 1.2% TBE agarose gel along with a 1 kb ladder of standards (Chimerx, Madison, WI). Gel samples were electrophoresed at 100 volts for approximately 2 hours and dried. Dried gels were autoradiographed at -70°C for 3 days and developed to visualize bands. The results are presented in Fig. 3, which presents autoradiographic data showing size-fractionated reverse transcriptase products using poly A-tailed mRNA as a template and oligo dT_{30} primers. In particular, the template was 891 nucleotides (lanes 2 and 4) or 7,500 nucleotides (lanes 1 and 3), nRT was used in reactions analyzed in lanes 3 and 4, while rRT was used in reactions analyzed in lanes 1 and 2. Both the native and recombinant RTs produced products of 891 bp and 7.5 kb, depending on the size of the mRNA template.

Example 7

The properties of native MAV-RT and recombinant RT were compared. In particular, optima for temperature, pH, magnesium ion concentration, and other divalent cation (*i.e.*, calcium, copper, manganese and zinc) concentrations were determined.

a) Temperature optima

The RNA-dependent DNA polymerase activity of native MAV-RT and recombinant MAV-RT (*i.e.*, M1BAhis) were compared in RT assays conducted at different temperatures.

The relative RT activities of the enzymes were compared between 37°C and 70°C at pH 8.0. The activity assays were performed in a 50 µl reaction mixture, containing 50 mM Tris HCl, pH 8, 40 mM KCl, 10 mM $MgCl_2$, 1.34% trehalose, 2% maltitol, 1 mM DTT, 0.5 mM poly rA:dT_{18}, 0.5 mCi [^{3}H]-dTTP (70-90 Ci/mmol), 0.5 mM dTTP, 5 U enzyme (rRT or nRT), and ddH_2O. Duplicate reactions were incubated at each temperature for 10 minutes. Products were quantitated by determining the [^{3}H]-dTTP incorporated using a scintillation counter. The results are presented as counts per minute as a function of temperature in degrees Celsius, as shown in Fig. 4A (black: rRT (*i.e.*, M1BAhis); hatched: nRT). These results reveal that the optimum temperature for both nRT and rRT in RT assays was 55°C.

The temperature profiles of nRT and rRT (*i.e.*, M1BAhis) in RAMP assays were also determined. RAMP reactions were conducted as described in PCT/US97/04170, incorporated herein by reference in its entirety. In particular, the target nucleic acid being amplified was Cryptosporidium mRNA from one oocyte. As described in detail in Example 8, this mRNA target was reverse transcribed into cDNA at different temperatures using 20 units of native RT or 15 units of recombinant RT. The results are presented absorbance at 450 nm as a function of temperature in degrees Celsius, as shown in Fig. 4C (hatched line: standard; closed circles: rRT (*i.e.*, M1BAhis)). The actual absorbance values at the various temperatures are shown below the figure (upper row: standard; lower row: rRT).

b) pH optima

The relative RT activities of nRT and rRT in reactions at various pH values were also compared. Two sets of comparative reactions were designed: one set incubated at a conventional temperature of 37°C, the other set incubated at a 60°C temperature suitable only for thermostable enzymes.

The pH values of selected buffers were adjusted at room temperature. Activity assays were performed in a 50 µl reaction mixture, containing 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.5 mM poly rA:dT_{18}, 0.5 mCi [^{3}H]-dTTP (70-90 Ci/mmol), 0.5 mM dTTP, 5 units enzyme, ddH_2O, and 50 mM Tris-HCl (pH 6, 7, 8, 8.3, 9, or 9.5). Reactions were incubated at 37°C or 60°C for 10 minutes. Products were quantitated by determining the [^{3}H]-dTTP incorporated as counts per minute using a scintillation counter.

with the activities of nRT and rRT (*i.e.*, M1BAhis) under the various pH conditions being shown in Figs. 4D and 4E (black: rRT (*i.e.*, M1BAhis); gray-hatched: nRT). The data in the Figures establish that the optimum pH for nRT and rRT is pH 8.

c) Mg^{++} ion optima

The RT assay described in Example 7(b) was modified to determine the influence of $MgCl_2$ concentration on the activities of the native and recombinant RTs. The reaction buffer contained 50 mM Tris-HCl, pH 8.3, and $MgCl_2$ ranging in concentration from 0-100 mM; all other reaction components were as described in Example 7(b). The reactions were incubated at 37°C. Incorporated [^{3}H]-dTTP was measured by scintillation counting, with the results presented as counts per minute. The optimum $MgCl_2$ concentration was found to be 5 mM for both nRT and rRT, as shown in Fig. 4F (black: rRT (*i.e.*, M1BAhis); gray-hatched: nRT).

d) Other divalent cation requirements

The reaction described above in the context of determining Mg^{++} concentration optima was modified to determine the influence of different divalent cations on RT activity. The reaction buffer included 50 mM Tris-HCl, pH 8.3, and 10 mM of the chloride salt of a divalent cation ($MgCl_2$, $CuCl_2$, $MnCl_2$, $ZnCl_2$, or $CaCl_2$). Independent experiments were performed and a curve was constructed. Fig. 4G shows the activities of the enzymes as counts incorporated as a function of the cation used in the reaction (black: rRT (*i.e.*, M1BAhis); gray-hatched: nRT). As shown in Fig. 4G, maximal activity of both nRT and rRT (*i.e.*, M1BAhis) was achieved using magnesium as the divalent cation.

Example 8

Conceptually, RT-PCR consists of a pre-amplification reaction followed by an amplification reaction. The pre-amplification reaction involves the use of reverse transcriptase to synthesize the first strand of cDNA using a CAT (*i.e.*, chloramphenicol acetyltransferase) mRNA as template. The CAT mRNA was provided in the Superscript kit from GIBCO-BRL, and the reaction was performed according to the supplier's recommendations. Following this reaction, the RNA from the RNA-DNA hybrid was removed by RNase H to free the first strand for use as a template in a Polymerase Chain Reaction (PCR).

The pre-amplification reaction mixture initially consisted of 50 ng of control mRNA (*i.e.*, CAT mRNA), 500 ng of oligo dT_{12-18}, and ddH_2O to bring the mixture to a total

volume of 12 μl. This mixture was incubated at 70°C for 1 minute. Subsequently, 2 μl of 10x PCR buffer, 2 μl of 25 mM $MgCl_2$, 1 μl of dNTP from a combined stock solution containing 10 mM each of dGTP, dATP, TTP and dCTP, and 2 μl of 0.1 M DTT were added to the mRNA/oligo dT mixture. One set of reactions was incubated with 20 U of nRT and the other set of reactions was incubated with 20 U of rRT (*i.e.*, M1BAhis). One tube from each set was incubated at one of several temperatures and each reaction proceeded for one hour. The reactions were terminated by incubation at 90°C for 2 minutes. Reactions were then cooled on ice and 1 μl of RNase H was added to each tube and incubated at 37°C for 20 minutes.

For the amplification reactions, each reaction mixture was assembled in a thin-wall tube containing: 5 μl of 10x PCR buffer, 3 μl of 25 mM $MgCl_2$, 1 μl of dNTP from a combined stock solution containing 10 mM each of dGTP, dATP, TTP and dCTP, 1 μl each of 10 μM amplification primer 1 and 10 μM amplification primer 2 as supplied in the superscript kit, 1 μl of Taq DNA polymerase and *Pyrococcus woesii* (*i.e.*, Pwo) DNA polymerase mix, (Boehringer Mannheim Corp., Indianapolis, IN) 2 μl of the cDNA mixture from the first-strand synthesis reaction and ddH_2O to 50 μl total volume. Reaction products were analyzed by subjecting 5 μl of the reaction to fractionation on a 1.2% TBE agarose gel and determining the intensity of the bands, in ng of DNA, using an imager equipped with a DC40 camera and Kodak Digital Sciences 1D™ software. The quantity of DNA synthesized by rRT was comparable to the quantity synthesized by nRT.

The results showed that the temperature optimum for RT-PCR was 60°C using either nRT or rRT, as shown in Fig. 4B (results are presented as ng of PCR products produced as a function of temperature in degrees Celsius, with open squares indicating rRT (*i.e.*, M1BAhis) and solid squares indicating nRT). The quantity of gene-specific products was greater at 60°C than at 37°C. The optimum temperature for RNA-dependent DNA polymerase activity for both nRT and rRT was 55°C (*see*, Example 7a and Fig. 4A). The differences in temperature optima are probably due to the need for both DNA-dependent DNA polymerase and RNase H activities (having different temperature optima) in RT-PCR.

Rapid Amplification (*i.e.*, RAMP) is an amplification technique disclosed in International Application Serial No. PCT/US97/04170. A RAMP reaction was also performed using an RT according to the invention and a first strand of cDNA from a *Cryptosporidium* oocyte mRNA as a template, along with a nicking enzyme (*i.e.*,

*Bst*HKCl) and Bst DNA polymerase. The Bst DNA polymerase provided both polynucleotide synthesis activity and strand displacing activity.

The reaction consisted of 35 mM K•PO_4, 0.7 mM Tris-HCl, pH 7.9, 1.4 mM dCTP, 0.5 mM each of dATP, dGTP and dTTP, 35 mg of Bovine Serum Albumin, 10.2 mM $MgCl_2$, 3.4 mM KCl, 0.7 mM DTT, 2% Maltitol, 1.34% Trehalose, 0.5 mM of Amplification Primer 1 (5′-ACCCCATCCAATGCATGTCTCGGGTCGTAGTCT-TAACCAT-3′; SEQ ID NO:31) and Amplification Primer 2, (5′-CGATTCCGCTC-CAGACTTCTCGGGTGCTGAAGGAGTAAGG-3′; SEQ ID NO:32) and 1% glycerol. To each reaction, 15 units of rRT (*i.e.*, M1BAhis) or 20 units of nRT were added, along with 36 units of Bst DNA polymerase and 250 units of *Bst*HKCl in a total volume of 10 μl.

The amount of product synthesized in each reaction was measured by a plate assay. The plate assay consisted of a gene-specific capture primer (5′-AAACTATGCCAACTAGAGATTGGAGGTTGTTT-3′; SEQ ID NO:30) bound to the wells of a microtiter plate and used to capture the product. The captured product was then detected by an oligonucleotide (HRP-conjugated P2 Comp; SEQ ID NO:37) linked to Horse Radish Peroxidase. The amount of bound HRP was detected by a colorimetric assay standard in the art.

The amount of product synthesized by the rRT was two-fold more than the quantity synthesized by nRT between temperatures of 55°C to 64°C, as shown in Fig. 4C. The difference in temperature optima between the RT assays and the amplifications may be due, in part, to the differences in the relative RNase H activities at the assessed temperatures. The lowest RNase H activity was seen between 60°-65°C, temperatures that also produced longer cDNA products and greater amplification of templates. The temperature profile of the RNase H activity of rRT is shown in Fig. 6B.

Example 9

In addition to RNA-dependent DNA polymerase activity, MAV-RT has additional enzyme activities, such as DNA-dependent DNA polymerase activity. The DNA-dependent DNA polymerase activity was investigated using a single-stranded M13mp18 DNA template and a sequence-specific [γ^{32}P] labeled primer (*i.e.*, Forward Sequencing Primer or FSP; 5′-CGCCAGGGTTTTCCCAGTCACGA-3′; SEQ ID NO:29). The 10 μl reaction mixture contained 50 mM Tris HCl, pH 8.3, 40 mM KCl, 10 mM $MgCl_2$, 20

μM of each conventional dNTP, 0.24 pmol of sequence-specific primer FSP and 800 ng of single-stranded M13mp18 DNA template. Four units of rRT (*i.e.*, M1BAhis) and 5 units of nRT were compared to a commercially available thermostable DNA polymerase (Sequitherm: 5 units) using the buffer provided in the kit. (Sequitherm Cycle Sequencing kit, Epicenter Technologies, Madison, WI). The DNA-dependent DNA polymerase activities of nRT and rRT were approximately equivalent.

The DNA-dependent DNA polymerase activity was also determined at different temperatures. For these reactions, incorporated [α-^{32}P]-dTTP served as a label and a non-radioactive primer was used. The reaction consisted of 200 ng of single-stranded M13mp18 DNA, 1-5 pmoles of FSP, 50 mM Tris-HCl, pH 8.3, 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.6 μCi of [α-^{32}P]-dTTP (3,000 Ci/mmol), and 20 μM each of dATP, dGTP, dCTP, and dTTP, in a total volume of 24 μl. A conventional protocol was used for the reactions (Sambrook *et al.*, (1989)) and the reactions were terminated by adding 2 μl of 10 mM EDTA (0.8 mM final concentration). The incorporated [α-^{32}P]-dTTP was determined using DE52 membranes and scintillation counting, as described above. Results shown in Fig. 5 indicate that the optimum temperature for DNA-dependent DNA polymerase activity for rRT was 45°C-50°C; for nRT, the temperature optimum was 55°C. The DNA-dependent DNA polymerase activities of the RTs of the invention broadens the range of applications amenable to use of these polypeptides. In addition to copying DNA as well as RNA, the enzymes may be used in any of the above-mentioned variety of amplification technologies known in the art. In addition, the polypeptides of the invention may be used to sequence RNA or DNA targets using Sanger's enzymatic approach as originally disclosed or any one of the many variations of that technique that have been developed since that time.

Example 10

An rRT (*i.e.*, M1BAhis) according to the invention (*i.e.*, M1BAhis) was subjected to an RNase H assay, using a protocol known in the art. Hillenbrand *et al.*, Nucl. Acids Res. 10:833 (1982). Reactions (25 μl) contained 20 mM HEPES-KOH, pH 8.0 (23°C), 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.24 mM [α-^{32}P] poly(A)-poly(dT) (1:2; 15 μCi/ml), and 4 μl of diluted enzyme purified from M1BAhis as described above.

For control reactions, standard stocks of RNase H (Molecular Biology Resources, Inc. Milwaukee, WI) with known activity were assayed in the range of 0.05 to 0.5

units/reaction (one unit of activity is defined as the amount of enzyme required to produce 1 nmol of acid soluble ribonucleotide from [α-^{32}P] poly A-poly(dT) in 20 minutes at 37°C). Two reactions were run without enzyme to serve as negative controls.

A reaction mixture, less enzyme, was prepared and the reaction started by the addition of enzyme. After 20 minutes of incubation at 37°C, the reaction was terminated by adding 25 μl of cold yeast tRNA as co-precipitant (10 mg/ml in 0.1 M sodium acetate, pH 5.0) followed by 200 μl of 10% trichloroacetic acid. Samples were then placed on ice for at least 10 minutes. The mixtures were centrifuged for 7 minutes at 16,000 x *g* in an Eppendorf microcentrifuge (Brinkman Instruments, Westburg, NY), and 200 μl of the supernatant fluid was withdrawn and counted in 5 ml of scintillation fluid.

The RNase H activity of the rRT at different temperatures was also tested using the reaction mixture described above. The results are presented as counts per minute of released radiolabeled ribonucleotide for each of two trials, as shown in Fig. 6A (black: rRT (*i.e.*, M1BAhis); gray-hatched: nRT). The data show that rRT had RNase H activity comparable to that of native RT. In addition, rRT activity was assessed at a variety of temperatures and the results presented in Fig. 6B showed that rRT was active over a wide range of temperatures. The optimum RNase H activity for rRT was 50°C. In contrast, RNase H activity was relatively low at temperatures of 37°C, 60°C and 65°C. Because of differences in the temperature optima for RT RNase H activity and the other RT activities, such as the RNA- and DNA-dependent DNA polymerase activities, the various methods relying on RT activity may be optimized by adjusting the temperature to achieve the desired mix of activities. For example, methods involving use of an RNase H activity may be performed at temperatures relatively close to the 55°C temperature optimum for the RNase H activity of rRT. Methods that benefit from decreased RNase H activity, such as RT-PCR and RAMP, may be performed at 60-65°C to maintain a low level of RNase H activity.

Example 11

A variety of polynucleotides were constructed that encoded modified RT fragments. These modified RTs include α and β polypeptides that have been terminally modified by deletion of a naturally occurring terminal region of the peptide to produce α-like and β-like fragments retaining RNA-dependent DNA polymerase activity. Other modified RTs according to the invention involve an α-like or β-like fragment attached at either the N-

terminus, C- terminus, or both termini to one or more peptides (those peptides including simple homo-oligomeric peptides, preferably charged or bulky, and peptides containing useful functionalities such as DNA binding, metal binding, structure stabilizing and polymerizing [*e.g.*, zinc finger domains, leucine zipper motifs, an NS1 binding site, GPRP (single-letter amino acid identification) or its inverse PRPG, among others] capacities). Yet other modified RTs according to the invention include fragments that lack a sequence found internally in one of the native polypeptides, α or β.

Techniques used to construct polynucleotides encoding these modified RTs are known in the art and described in Examples 1 and 3 above. Generally, the strategy was to use PCR to construct the desired polynucleotide, which was then cloned and expressed to produce the encoded modified RT. The expression studies were generally conducted as described in Example 4.

Expression of eukaryotic genes in prokaryotes may result in production of misincorporated, truncated and/or insoluble proteins (misfolding) due to the presence of rare codons in those eukaryotic genes. Translation of these rare codons is limited by the regulated expression of tRNAs corresponding to these rare codons. Hence, expression of eukaryotic genes having abundant rare codons sometimes results in misincorporation, truncation and/or misfolding. One approach to minimizing such problems is to clone the tRNA corresponding to these rare codons and express the clone in *E.coli* in order to facilitate the expression of eukaryotic genes. We have cloned and expressed the ArgU tRNA because the arginine codons (AGG, AGA CGA and CGG) present the largest number of rare codons in AMV-RT. Co-expression of AMV-RT and ArgU is expected to improve expression (*i.e.*, activity levels) of AMV-RT. Other rare codons such as leucine (CTA) and proline (CCC) will also be cloned and co-expressed.

Another approach to improved expression of the modified RTs of the invention in prokaryotes is to change the rare codons in modified RT coding regions to frequently used codons. Such changes can be readily effected by a variety of techniques known in the art, *e.g.*, site-directed mutagenesis using synthetic oligonucleotides. In an *E. coli* expression system, there would be 90 rare codons (38 arginine, 23 proline, 15 isoleucine, 10 leucine and 4 serine codons) in the AMV-RT gene, all or some of which may be advantageously changed to frequent codons. Changing all 90 rare codons to the frequent codons found in abundantly expressed genes could imbalance host cell metabolism, however. To accommodate deleterious effects on host cell metabolism arising from modified RT expression levels that are

too high, a library of clones may be constructed using, e.g., an M13-based approach to site-directed mutagenesis involving oligonucleotide primer incorporation. Specifically, pools of synthetic oligonucleotides, each oligonucleotide designed to convert one or a few rare codons to frequent codons, and a template comprising a modified RT coding region may be used to synthesize a collection of modified RTs having a range of 1-90 rare codon conversions. Clones having RT activity may be isolated from this library by conventional screening techniques (e.g., binding to radioactive substrate and activity assays, among others).

To facilitate an understanding of the structures of the various polynucleotides and polypeptides disclosed in this Example, Table II below collects pertinent information. All constructions generated by PCR used a suitable, full-length coding region sequence as a template, such as the *pol* gene sequence found in M1-5,6.

Table II

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BA (His6)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BAhisXhoI extend (SEQ ID NO:45)	253-269 1986-1967	6 His codons
M1BA (His10)	1766	FM1BA SmaI (SEQ ID NO:25); RM1BA His10 (SEQ ID NO:46)	253-269 1986-1967	10 His codons
M1BA (His12)	1772	FM1BA SmaI (SEQ ID NO:25); RM1BA His12 (SEQ ID NO:47)	253-269 1986-1967	12 His codons
M1BA (Leu)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA Leu (SEQ ID NO:48)	253-269 1986-1968	6 Leu codons

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BA (Lys)	1757	FM1BA SmaI (SEQ ID NO:25); RM1BA Lys (SEQ ID NO:49)	253-269 1986-1968	7 Lys codons
M1BA (Arg6)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA Arg6 (SEQ ID NO:50)	253-269 1986-1967	6 Arg codons
M1BA (Arg3, X4)	1757	FM1BA SmaI (SEQ ID NO:25); RM1BA Arg3X4 (SEQ ID NO:51)	253-269 1986-1967	3 Arg, 2 Asn, 1 Gln, 1 Tyr codon
M1BA (Asp6)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA Asp6 (SEQ ID NO:52)	253-269 1986-1968	6 Asp codons
M1BA (Asp4)	1748	FM1BA SmaI (SEQ ID NO:25); RM1BA Asp4 (SEQ ID NO:53)	253-269 1986-1968	4 Asp codons
M1BA (Asp5)	1751	FM1BA SmaI (SEQ ID NO:25); RM1BA Asp5 (SEQ ID NO:54)	253-269 1986-1968	5 Asp codons
M1BA (Asp8)	1760	FM1BA SmaI (SEQ ID NO:25); RM1BA Asp8 (SEQ ID NO:55)	253-269 1986-1968	8 Asp codons

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BA (Asp12)	1772	FM1BA SmaI (SEQ ID NO:25); RM1BA Asp12 (SEQ ID NO:56)	253-269 1986-1968	12 Asp codons
M1BA (Glu6, XhoI)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA Glu6, XhoI (SEQ ID NO:57)	253-269 1986-1968	6 Glu codons
M1BA (Glu12)	1772	FM1BA SmaI (SEQ ID NO:25); RM1BA Glu12 (SEQ ID NO:58)	253-269 1986-1968	12 Glu codons
M1BK 620	1862	FM1BA SmaI (SEQ ID NO:25); RM1BK 620 (SEQ ID NO:74)	253-269 2112-2092	
M1BK 620 His	1880	FM1BA SmaI (SEQ ID NO:25); RM1BK 620 His (SEQ ID NO:60)	253-269 2112-2092	6 His codons
M1BK 640 XhoI	1919	FM1BA SmaI (SEQ ID NO.25); RM1BK 640 XhoI (SEQ ID NO:76)	253-269 2149-2169	
M1BK 660 XhoI	1982	FM1BA SmaI (SEQ ID NO:25); RM1BK 660 XhoI (SEQ ID NO:77)	253-269 2210-2232	

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BK680 XhoI	2042	FM1BA SmaI (SEQ ID NO:25); RM1BK 680 XhoI (SEQ ID NO:78)	253-269 2273-2292	
M1BK 760 XhoI	2282	FM1BA SmaI (SEQ ID NO:25); RM1BK 760 XhoI (SEQ ID NO:79)	253-269 2512-2532	
M1BK 800 XhoI	2399	FM1BA SmaI (SEQ ID NO:25); RM1BK 800 XhoI (SEQ ID NO:80)	253-269 2628-2649	
M1BK 640 His XhoI	1937	FM1BA SmaI (SEQ ID NO:25); RM1BK 640 His XhoI (SEQ ID NO:81)	253-269 2149-2169	6 His codons
M1BK 660 His XhoI	2000	FM1BA SmaI (SEQ ID NO:25); RM1BK 660 His XhoI (SEQ ID NO:82)	253-269 2210-2232	6 His codons
M1BK 680 His XhoI	2060	FM1BA SmaI (SEQ ID NO:25); RM1BK 680 His XhoI (SEQ ID NO:83)	253-269 2273-2292	6 His codons

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BK 760 His XhoI	2300	FM1BA SmaI (SEQ ID NO:25); RM1BK 760 His XhoI (SEQ ID NO:100)	253-269 2512-2532	6 His codons
M1BK 800 His XhoI	2417	FM1BA SmaI (SEQ ID NO:25); RM1BK 800 His XhoI (SEQ ID NO:84)	253-269 2628-2649	6 His codons
M1BA (LZIP2 XhoI)	1757	FM1BA SmaI (SEQ ID NO:25); RM1BA LZIP2 XhoI (SEQ ID NO:61)	253-269 1986-1968	Leucine zipper (2 copies)
M1BA (LZIP3 XhoI)	1778	FM1BA SmaI (SEQ ID NO:25); RM1BA LZIP3 XhoI (SEQ ID NO:62)	253-269 1986-1968	Leucine zipper (3 copies)
M1BA (LZIP4 XhoI)	1799	FM1BA SmaI (SEQ ID NO:25); RM1BA LZIP4 XhoI (SEQ ID NO:63)	253-269 1986-1968*	Leucine zipper (4 copies)
M1BA (LZIP5 XhoI)	1820	FM1BA SmaI (SEQ ID NO:25); RM1BA LZIP5 XhoI (SEQ ID NO:64)	253-269 1986-1968*	Leucine zipper (5 copies)

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BA (Cyst2)	1742	FM1BA SmaI (SEQ ID NO:25); RM1BA Cyst2 (SEQ ID NO:65)	253-269 1986-1968	2 Cys codons
M1BA (Cyst6)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA Cyst6 (SEQ ID NO:66)	253-269 1986-1968	6 Cys codons
M1BA (GPRP)	1748	FM1BA SmaI (SEQ ID NO:25); RM1BA GPRP (SEQ ID NO:67)	253-269 1986-1968	GPRP motif
M1BA (PRPG)	1748	FM1BA SmaI (SEQ ID NO:25); RM1BA PRPG (SEQ ID NO:68)	253-269 1986-1968	PRPG motif
M1BA (NS1 XhoI)	1796	FM1BA SmaI (SEQ ID NO:25); RM1BA NS1 XhoI (SEQ ID NO:98)	253-269 1986-1966	NS1 site
M1BA (WH)	1769	FM1BA SmaI (SEQ ID NO:25); RM1BA WH (SEQ ID NO:69)	253-269 1986-1968	WH motif
M1BA (3PPG XhoI)	1763	FM1BA SmaI (SEQ ID NO:25); RM1BA 3PPG XhoI (SEQ ID NO:70)	253-269 1986-1968	3 "PPG" motifs

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
M1BA (Trp)	1754	FM1BA SmaI (SEQ ID NO:25); RM1BA TRP (SEQ ID NO:71)	253-269 1986-1968	6 Trp codons
M1BA (Nhis SmaI)	1754	FM1BA Nhis SmaI (SEQ ID NO:72); RM1BA XhoI (SEQ ID NO:59)	253-269 1986-1967	6 His codons
M1BA (NWH SmaI)	1769	FM1BA NWH SmaI (SEQ ID NO:73); RM1BA XhoI (SEQ ID NO:59)	253-270 1986-1967	WH motif
DNPCR1 (D450N)	1754	FDNPCR1 (D450N) (SEQ ID NO:92); RDNPCR1 (D450N) (SEQ ID NO:93)	1577-1622	Mismatch at position 1600 of SEQ ID NO:1
DNPCR2 (D505N)	1754	FDNPCR2 (D505N) (SEQ ID NO:94); RDNPCR2 (D505N) (SEQ ID NO:95)	1744-1789	mismatch at position 1765 of SEQ ID NO:1
M1BA (E484Q)	1754	FM1BA E484Q (SEQ ID NO:96); RM1BA E484Q (SEQ ID NO:97)	1678-1725	mismatch at position 1702 of SEQ ID NO:1

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
Core domain deletion-Fragment 1a	2113	FM1BA SmaI (SEQ ID NO:25); and RM1BK 620 XhoI (SEQ ID NO:74); and F Cint XhoI (SEQ ID NO:85); R Cint SalI (SEQ ID NO:86)	253-269 2092-2112 2560-2580 2788-2811	
Core domain deletion-Fragment 1b	2170	FM1BA SmaI (SEQ ID NO:25); and RM1BK 640 XhoI; and F Cint XhoI (SEQ ID NO:85); R Cint SalI (SEQ ID NO:86)	253-269 2149-2169 2560-2580 2788-2811	
Core domain deletion-Fragment 1c	2233	FM1BA SmaI (SEQ ID NO:25); and RM1BK 660 XhoI; and F Cint XhoI (SEQ ID NO:85); R Cint SalI (SEQ ID NO:86)	253-269 2210-2232 2560-2580 2788-2811	
Core domain deletion- 3' fragment 2a	2131	FM1BA SmaI (SEQ ID NO:25); and RM1BK 620 XhoI (SEQ ID NO:74); and F Cint XhoI (SEQ ID NO:85); R Cint His SalI (SEQ ID NO:87)	253-269 2092-2112 2560-2580 2788-2811	6 His codons

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
Core domain deletion- 3' fragment 2b	2188	FM1BA SmaI (SEQ ID NO:25); and RM1BK 640 XhoI; and F Cint XhoI (SEQ ID NO:85); R Cint His SalI (SEQ ID NO:87)	253-269 2149-2169 2560-2580 2788-2811	6 His codons
Core domain deletion- 3' fragment 2c	2251	FM1BA SmaI (SEQ ID NO:25); and RM1BK 660 XhoI; and F Cint XhoI (SEQ ID NO:85); R Cint His SalI (SEQ ID NO:87)	253-269 2210-2232 2560-2580 2788-2811	6 His codons
Core domain deletion- 3' fragment 3a	2155	FM1BA SmaI (SEQ ID NO:25); and RM1BK 620 XhoI (SEQ ID NO:74); and F Cint 731 SalI (SEQ ID NO:88); RCint 830 XhoI (SEQ ID NO:90)	253-269 2092-2112 2443-2463 2736-2716	
Core domain deletion- 3' fragment 3b		FM1BA SmaI (SEQ ID NO:25); and RM1BK 640 XhoI; and F Cint 731 SalI (SEQ ID NO:88); RCint 830 XhoI (SEQ ID NO:90)	253-269 2149-2169 2443-2463 2736-2716	

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
Core domain deletion- 3' fragment 3c	2275	FM1BA SmaI (SEQ ID NO:25); and RM1BK 660 XhoI; and F Cint 731 SalI (SEQ ID NO:88); RCint 830 XhoI (SEQ ID NO:90)	253-269 2210-2232 2443-2463 2736-2716	
Core domain deletion- 3' fragment 4a	2101	FM1BA SmaI (SEQ ID NO:25); and RM1BK 620 XhoI (SEQ ID NO:74); and F Cint 751 SalI (SEQ ID NO:89); RCint 830 XhoI (SEQ ID NO:90)	253-269 2092-2112 2497-2517 2736-2716	
Core domain deletion- 3' fragment 4b	2158	FM1BA SmaI (SEQ ID NO:25); and RM1BK 640 XhoI; and F Cint 751 SalI (SEQ ID NO:89); RCint 830 XhoI (SEQ ID NO:90)	253-269 2149-2169 2497-2517 2736-2716	
Core domain deletion- 3' fragment 4c	2221	FM1BA SmaI (SEQ ID NO:25); and RM1BK 660 XhoI; and F Cint 751 SalI (SEQ ID NO:89); RCint 830 XhoI (SEQ ID NO:90)	253-269 2210-2232 2497-2517 2736-2716	

Clone	Approx. Length of Coding Region	Oligonucleotide additions/PCR Primers	Added oligonucleotide locations (nucleotide numbering of SEQ ID NO:1)	Added oligonucleotide characteristics
Core domain deletion- 3' fragment 5a	2032	FM1BA SmaI (SEQ ID NO:25); and RM1BK 620 XhoI (SEQ ID NO:74); and F Cint 771 SalI (SEQ ID NO:99); RCint 830 XhoI (SEQ ID NO:90)	253-269 2092-2112 2566-2586 2736-2716	
Core domain deletion- 3' fragment 5b	2089	FM1BA SmaI (SEQ ID NO:25); and RM1BK 640 XhoI; and F Cint 771 SalI (SEQ ID NO:99); RCint 830 XhoI (SEQ ID NO:90)	253-269 2149-2169 2566-2586 2736-2716	
Core domain deletion- 3' fragment 5c	2152	FM1BA SmaI (SEQ ID NO:25); and RM1BK 660 XhoI; and F Cint 771 SalI (SEQ ID NO:99); RCint 830 XhoI (SEQ ID NO:90)	253-269 2210-2232 2566-2586 2736-2716	

' Oligonucleotides hybridize to an internal region of oligonucleotide RM1BA LZip3 XhoI, which in turn recognizes the indicated region of SEQ ID NO:1.

A. Terminally deleted RTs

The full-length RT coding region was truncated by deletions using conventional methodologies described above (*e.g.*, Example 3). One set of deletion derivatives lacked the 3' end of the MAV-RT coding region to varying extents. Again, relative to the full-length gene (SEQ ID NO:1), the 3' (C-terminal) deletion extending to the *KpnI* site (MIKA; see SEQ ID NO:8) increased the RT expression level, as evidenced by SDS-

PAGE. Relative to the full-length gene (SEQ ID NO:1), deletion of the region extending from the *Bgl*II site to the 3′ terminus (M1BA; see SEQ ID NO:6) also increased RT expression and activity, as evidenced by SDS-PAGE and activity assays (see below). The C-terminally truncated RTs (M1KA and M1BA) have lengths that fall in between the lengths of the native α and β polypeptides. Relative to the alpha fragment of MAV-RT, the beta fragment has an additional 254 amino acids at the C-terminus, which provides an integrase activity. This region of the polypeptide contributes to the insolubility of the polypeptide and reduces its recovery from cell extracts, as shown by the relative insolubility of a (+) integrase form of RT (*e.g.*, the M1KA gene product, see below) compared to a (-) integrase form (*e.g.*, the M1BA gene product). Because the integrase domain is only needed for the retroviral life cycle and not for the RNA- or DNA-dependent DNA polymerase activities, this region was deleted in M1BA (α-like fragment). Note that the M1BA α-like fragment (amino acids 1-578 of SEQ ID NO:2) is larger than the naturally occurring α fragment of MAV-RT (amino acids 1-573 of SEQ ID NO:2). Without wishing to be bound by theory, this deletion was expected to result in an increase in the solubility, and hence recovery, of the protein.

A series of clones was constructed to express the M1BA and M1KA series of modified RTs, which have C-terminal deletions in order to increase the levels of expression and to stabilize the RT activity (RNA-dependent DNA polymerase activity). Convenient restriction sites in full-length clones such as PMBacRT and pHRT, *e.g.*, *Bgl* II (spanning nucleotides 1,986-1,991 of SEQ ID NO:1) and *Kpn*I (spanning nucleotides 2,745-2,750 of SEQ ID NO:1), were used to eliminate the 3′ end of the coding region of the RT gene (see, Table I). The 3′ deletion derivatives, encoding RT polypeptide fragments having C-terminal deletions, were obtained by *Bgl*II-*Pst*I or *Kpn*I-*Pst*I restrictions of pMBacRT and pHRT, respectively (*Bgl*II and *Kpn*I sites in the MAV-RT coding region; *Pst*I site in the vector). Recombinant molecules containing the *Bgl II-Pst*I 3′ terminal deletion were designated pBacMIBA and pHBRT (pH33ΔBP6) and recombinant molecules containing the *Kpn*I-*Pst*I deletion were designated pBacMIKA and pHKRT (pH33ΔKP5). The deletion derivatives pBacMIBA and pBacMIKA had approximately 1.17 and 0.4 kb deletions from the 3′ end of the full-length gene (see, SEQ ID NO:1), respectively. The fragment bounded at its 3′ end by the *Bgl*II site (SEQ ID NO:6) was used to express an alpha-like RT fragment (the α-like fragment, M1BA, contained amino acids 1-578 of SEQ ID NO:2; native MAV-RT α contains amino acids 1-572 of SEQ ID NO:2) and the

fragment bounded by the *Kpn*I site (SEQ ID NO:8) was used to express a beta-like RT fragment (the β-like fragment, M1KA, contained amino acids 1-832 of SEQ ID NO:2; native MAV-RT β contains amino acids 1-858 of SEQ ID NO 2).

Miniprep and sequencing analyses were done to confirm the identities of the recombinant clones described above. Recombinant viruses obtained from co-transfection with virus BacPak6 and transfer vector pBacM1BA or pBacM1KA were called M1BA and M1KA, respectively.

B. Alpha-like recombinants encoding non-native terminal peptides

1 Simple peptide tags

One category of α fragment modifications was designed to mimic one or more properties of the integrase domain found in the β fragment but missing from the α fragment of Type III RTs. Partial mimicking of the integrase domain, without the deleterious impact on solubility and host cell viability associated with the native integrase domain, was accomplished by adding polynucleotide sequences encoding His tags at the 3′ ends of the modified RT coding regions.

A His-tag addition to the C-terminus of an RT polypeptide was achieved by recombinant expression of a polynucleotide containing an RT coding region fused in-frame to His codons. In particular, the fusions were constructed by adding oligonucleotides containing 6 histidine codons to the 3′ end of the RT gene using ligase, as in the case of the construction of pBacM1KAhis, or by PCR amplification with oligonucleotides that specified 6 histidine codons, as in the case of the construction of pBacM1BAhis.

The basic nature of the added His amino acids was expected to increase binding to the negatively charged nucleic acids, enhancing the stability of the polypeptides. The increased stability, in turn, was expected to result in increased activity of amino-acid-tagged RTs, relative to their untagged counterparts. In addition, the His tags were expected to chelate metal ions (e.g., Ni^{++}), thereby potentiating polymerization of the modified RTs. A His-tagged RT (M1BAhis) was found in homo-polymeric form (molecular weight greater than 200 kDa), as determined using non-denaturing PAGE and molecular sieve chromatography with Superose 12HR10/30 (separation range of 1-300 kDa; Pharmacia-Upjohn).

Expression levels of the RT fragments modified by amino acid tagging showed that the structurally unstable alpha fragment was stabilized by addition of peptide tags to the C-terminus of the AMV-RT alpha fragment.

Other modified RTs bearing peptides at the C-terminus of the α-like fragment were generated by PCR, as described above. The forward and reverse PCR primers had codons corresponding to the - and C-termini of the AMV-RT alpha fragment, along with codons corresponding to the peptide tags to be added. A linearized template (pHSEM1) containing the full-length RT gene was used for the PCR amplifications. Additional information concerning this class of modified RTs, as well as the polynucleotides encoding them, is found in Table II

The PCR product was restricted with a suitable restriction enzyme and ligated to pBacPak9 that was digested with a compatible enzyme. The selected recombinants were sequenced to confirm addition of the appropriate tags.

2 C-terminal peptides exhibiting DNA binding properties

DNA binding motifs of proteins, may have either general affinity (*i.e.*, non-specific binding) or sequence-specific affinity for DNA. Several nucleic acid binding domains have been identified and reported to play a role in important cellular functions such as viral packaging, transcriptional and translational regulation, transport between the nucleus and cytoplasm, splicing, and stability, among others. Karaya et al., J. Biol. Chem. 266:11621-11627 (1991), Burd, et al., Science 265:615-621 (1994); Weiss, et al., Biopolymers 48(2-3) 167-180 (1998), Nassal, M., J. Virol. 66(7):4107-16 (1992) Ritt, et al., Biochemistry 37 2673-81 (1998) DNA binding domains with general affinity are preferable to target-specific binding domains because of the reduced substrate specificity of modified RTs having such general binding domains.

Several basic amino acids are known to enhance the affinity of a protein for nucleic acid templates. The positive charges of arginine, lysine, and histidine increase the non-specific affinity of polypeptides containing such residues for nucleic acid, thereby facilitating the search for specific binding sites. Several arginine-rich motifs and arginine-lysine-rich motifs have been identified in nucleic acid binding domains. The arginine-lysine rich motif ELKIKRLRKKFAQKMLRKARRK is involved in RNA binding, which could enhance the activity of RT. In addition, a lysine-rich protein is associated with DNA in the kinetoplast and plays a role in segregation of the kinetoplast DNA. Hines, Mol. and Biochem. Parasitol. 94 41-52 (1998) Similarly, acidic amino acid tags are reported to be involved in packaging of viral DNA. The packaging may be mediated through metal ions that have affinity for DNA. International Patent Publication No. WO 98/07869. Additionally, charged amino acids are

present on the surface of structural proteins and may play a role in stabilizing secondary structures.

The addition of histidine, glutamic acid, and aspartic acid tags enhanced the activity of the alpha fragment 20-100 fold. A peptide tag consisting of six arginine residues improved the activity five-fold. However, specific arginine-rich motifs such as RNRNRQY (Arg3X4, found at the C-terminus of the GP67 envelope glycoprotein proposed to be involved in baculoviral DNA packaging) enhanced (*i.e.*, increased or prolonged) activity by 20- to 40-fold. Other RNA- and DNA-binding motifs such as RRRDRGRS are expected to yield similar results. However, six continuous lysine residues did not increase the activity. A higher number of lysine residues or correct spacing of the lysine residues may be required for enhancement of function.

The mechanism of enhancement of activity due to these tags could be due to the increased structural stability of the recombinant or stability resulting from direct or metal-mediated nucleic acid binding.

M1BA	2000-3000 Units/g of insect cells
M1BA his	50000-200,000 U/g
M1BA arg6	15,750 U/g
M1BA lys6	2050 U/g
M1BA Arg3X4	57,000 U/g
M1BA glu6	170,000 U/g
M1BA asp6	40,000 U/g
M1BA leu6	2250-3900 U/g
Nhis M1BA asp4	95,000 U/g
Nhis M1BA asp5	115,250 U/g
Nhis M1BA asp6	236,250 U/g

Most of the sequence-specific DNA binding proteins have a general basic region and a sequence-specific region for binding to DNA. There are several sequence-specific DNA binding motifs such as zinc-finger domains (*e.g.*, TFIIIA CX2CX12HX3H) and the basic region of the bZIP family of proteins. Similarly, there are arginine-rich domains such as TRQARRNRRRWRARQR and YGRKKRRQRRRP that recognize specific RNA sequences that are also expected to enhance the activity of RT. The N-terminus of the RT integrase domain has a zinc-finger-like (Hx3HX23CX2C) motif. This N-terminus binds zinc and has been reported to both induce proper folding of the N-terminus, to be remarkably thermostable

as well. Burke et al., J. Biol. Chem. 267 9639-44 (1992). Because the full-length MAV-RT gene has a zinc-finger-like domain, the reverse primer used in some PCR amplifications included this region of the integrase (see Table II).

A beta-like derivative (620 amino acids) containing the zinc-finger-like motif was more active than the non-tagged alpha fragment (578 amino acids) and expressed 30,000 units per gram of cell pellet.

M1BK620	31,950 U/g
M1BK620 his	50,000-140,000 U/g

The addition of the sequence-specific, zinc-finger-like motif produced a lower level of RT activity than the His-tagged fragment, however. These results suggest that a general nucleic acid binding domain (His tag) may enhance RT activity to a greater extent than a sequence-specific domain (zinc-finger-like motif) and, therefore, could replace the sequence-specific zinc-finger-like motif of RT, leading to an increase in activity. General nucleic acid binding domains enhance the stability of both the 578- and the 620-amino-acid-length fragments.

3 C-terminal peptide tags having polymerization domains

Disulfide bond-forming domains (i.e., cysteine-rich regions) present in immunoglobulin genes are involved in disulfide bond formation between the light and heavy chains. Hence, addition of two cysteine residues at the C-terminus was anticipated to promote dimer formation through disulfide bonding.

Addition of two cysteine residues enhanced the activity of the alpha-like fragment; however, 6 contiguous cysteine residues reduced the activity of the modified RT.

M1BA	2000-3000 U/g
M1BA cyst2	190,000 U/g
M1BA cyst6	720 U/g

The GPRP (fibrin clotting) tetrapeptide is the primary polymerization pocket of the blood clotting protein fibrin. This domain is exposed at the amino terminus of fibrin monomers by proteolytic cleavage of the precursor protein. The domain then polymerizes by binding to complementary binding sites on other fibrinogen molecules to form clots. Because peptides were being added to the C-terminus of α-like constructs, the reverse-sequence tetrapeptide, PRPG, was also examined.

Addition of GPRP enhanced the RT activity approximately 50-fold, while addition of PRPG enhanced the activity of RT by approximately 100-fold. In other embodiments, the D-

isomers of amino acids are used in peptide tags. For example, D-isomers are used in generating PRPG peptides for use in preparing modified RTs of the invention.

M1BA	2000-3000 U/g
M1BA GPRP	107,500 U/g
M1BA PRPG	243,500 U/g

Histidine residues can also promote dimer formation mediated by metal ions. The addition of 6 His residues to the C-terminus of the α-like RT resulted in a 20- to 40-fold increase in activity. Additions of different length histidine tags are contemplated.

M1BA	2000-3000 U/g
M1BA his	50000-200000 U/g

NS1 is a DNA-binding protein produced by the minute virus of mice. The protein has replicational and transcriptional functions. Homo-oligomerization of NS1 is required for its function and a small region, N-VETTVTTAQETKRGRIQTK-C, of NS1 has been identified as the domain involved in oligomerization. Pujol et al., J. Virol 71:7393-7403 (1997). Addition of this peptide tag to the C-terminus of AMV-RT fragments enhanced RT activity.

M1BA	2000-3000 U/g
M1BA NS1	380,000 U/g

4 C-terminal peptide tags having metal binding domains

Histidine tags can be used as metal binding domains, as explained above. In addition, modified RTs having C-terminal His tags were constructed and subjected to expression analyses. The results, presented above, indicate that peptide tags, having metal binding capacity, enhance RT expression.

Zinc fingers also exhibit metal binding capacity and are also involved in DNA binding. As described above, the N-terminus of the integrase domain of MAV-RT has a zinc-finger-like (HX3HX23CX2C) motif. This N-terminus binds zinc and has been reported to induce proper folding of the N-terminus. It is expected that peptide tags containing one or more zinc-finger-like domains will enhance the activity of modified RTs in which they are found.

5 C-terminal peptide tags having structure-stabilizing domains

Other embodiments of the invention involve the addition of domains designed to structurally stabilize the alpha-like fragment so that it no longer requires a second fragment for structural stability. There are several motifs that have been identified and shown to form specific structures, such as alpha helices, beta sheets, and coils, among others, all of which are known in the art. Formation of defined structures facilitates the formation of active domains

and promotes interactions with other such domains. Beta strands and beta sheets frequently promote aggregation in, and precipitation from, solution. Desjarlais et al. Curr. Opin. in Biotechnol. 6:460-466 (1995). Hence, most of the C-terminal tag additions were capable of forming helices or coils. These secondary structure predictions are based on the well-known Chou and Fassman algorithms.

The WEAAH (WH) motif, comprising histidine and tryptophan, promotes formation of alpha helices, or defined structures, thereby giving structural stability to the protein.

M1BA	2000-3000 U/g
M1BA WH	104,720 U/g

Addition of the WH domain may extend the helix at the C-terminus and thereby enhancing the stability of the alpha fragment. Regardless of the reason, however, modified RTs containing a WH motif exhibit enhanced RT activity.

The "PPG" triple-helical domain is responsible for binding interactions in the structural protein collagen. This motif is responsible for the structural stability and proper assembly of collagen. Addition of peptides containing this motif in generating modified RTs according to the invention is expected to enhance the activity of such RTs relative to corresponding RTs lacking such peptides.

Addition of tryptophan residues is predicted to extend the α-helix at the C-terminus and to enhance the stability of the alpha-like fragment. Tryptophan is a bulky amino acid and could substitute for histidine tags in providing structural stability. Comparative assays showed that a domain comprising Trp residues enhanced RT activity approximately 50-fold.

M1BA	2000-3000 U/g
M1BA Trp	96,500 U/g

The GPRP and PRPG motifs identified in fibrin as the domains involved in interaction with other clotting proteins enhance the activity of the AMV-RT alpha-like fragment. This motif is predicted to form coil-turn-coil structures.

M1BA	2000-3000 U/g
M1BA GPRP	107,500 U/g
M1BA PRPG	243,500 U/g

The NS1 domain primarily forms beta sheets and coils. The presence of hydrophobic residues alone is not very desirable because they form beta sheets and are typically buried in the secondary structure of the protein. This may affect the natural folding of domains. Hence,

a motif that had a mixture of coils and beta sheets was chosen for analysis. Addition of this domain produced an active α-like fragment that appeared to be stable

M1BA	2000-3000 U/g
M1BA NS1	380,000 U/g

The leucine zipper motif is a helix-turn-helix motif which has been reported to dimerize by a coiled-coil interaction. This defined structure of the leucine zipper is expected to enhance the stability of the alpha-like fragment in addition to providing dimerization abilities.

M1BA	2000-3000 U/g
M1BA Lzip23	7170 U/g
M1BA Lzip3	1620 U/g

Addition of a single heptad repeat enhanced the activity by 2-3 fold. Addition of two heptad repeats did not improve the activity. However, additions of 4-5 heptad repeats produced RTs that had reduced activity levels.

6 N-terminal peptide tags

Consistent with the description in Examples 3 and 4 of N-terminal peptide tags being added to modified RTs that exhibited enhanced expression, several constructs were generated and characterized. One modified RT, NhisM1BA, contained a His tag attached to the N-terminus of an α-like fragment. Other RTs were modified to contain peptide tags at both termini (Nhis M1BA asp 4, Nhis M1BA asp 5, Nhis M1BA asp 6, and Nhis M1BA WH). Expression studies conducted as described in Example 4 led to the results shown below.

Nhis M1BA	10,000-41,700 U/g
M1BAChis	50,000-200,000 U/g
Nhis M1BA asp 4	95,000 U/g
Nhis M1BA asp 5	115,250 U/g
Nhis M1BA asp 6	236,250 U/g
Nhis M1BA WH	86,000 U/g

Expression of M1BAChis was measured to provide a relative control for the measurement of Nhis M1BA expression. The results show that activity of RTs modified by a His tag present at either the N-terminus or the C-terminus is increased relative to untagged RTs. Other variations, such as the addition of peptide tags to both termini of an RT (*e.g.*, an N-terminal His tag coupled to a C-terminal Asp-, Glu-, or Trp-His- (*i.e.*, WH) tag), are also contemplated by the invention. Large-scale expression studies have shown that similar activity levels of

approximately 100,000 units/g insect cells are achieved with M1BA asp (N-terminally modified RT) and Nhis M1BA asp (RT having 6 His residues at the N-terminus and 4-6 Asp residues at the C-terminus)

7 Peptide tagging of other Type III RTs

The strategies described above were also used to modify RTs from other avian sources, such as Rous Sarcoma Virus and Avian Tumor Virus. The C-terminal addition of a six-histidine peptide tag to an alpha fragment of each of these avian RTs substantially increased the RT activity, relative to the non-tagged AMV-RT α-like fragment.

M1BA	2000-3000 U/g
RSV-RT	43,350 U/g
ATV-RT	71,900 U/g

Therefore, the modification strategies applied to AMV-RT polynucleotides and polypeptides are applicable more generally to dimeric (*i.e.*, Type II and Type III) reverse transcriptase coding regions and polypeptides, and all of these modified RTs fall within the scope of the present invention.

C. Beta-like recombinants

Modifications of β RT

Polynucleotides encoding a variety of beta-like modified RTs were constructed using the techniques described in Example 3 and expressed using the techniques described in Example 4, along with M1-5,6 encoding the full-length AMV-RT. Expression of the full-length beta fragment resulted in low levels of highly insoluble, full-length protein, in both a eukaryotic (insect cell) and a prokaryotic (*E. coli*) host. Because expression of the full-length beta fragment resulted in mostly insoluble protein, the native beta polypeptide was modified in an effort to increase its solubility and, hence, activity. One strategy for modifying the β fragment involved deletions of parts of the native β RT. The native beta coding region specifies 858 amino acids and the full-length β-like fragment disclosed herein consists of 832 amino acids. Thus, the β-like fragment lacks the 26 C-terminal amino acids of full-length native β. Expression of the full-length β-like polypeptide, relative to the full-length native β, showed an increase of one-hundred-fold in expression, as evidenced by SDS-PAGE analysis; however, the β-like polypeptide was still highly insoluble (approximately 90% insoluble), resulting in a five-fold increase in activity.

M1KA	1000 U/Liter of cells
M1KAhis	2,000 U/L

M1-5 200 U/L

Modified RTs having C-termini between 580 and 832 amino acids (see SEQ ID NO:2) are also contemplated by the invention. Because both the 580- and the 620-amino-acid recombinants are soluble, and the 832- and 858-amino-acid recombinants are relatively insoluble, deletions that truncate the C-terminus to a position between 580-832 amino acids are expected to result in modified β-like polypeptides that are soluble. In particular embodiments, the β-like polypeptide has a C-terminus at any one of positions 580-832, such as positions 620, 640, 660, 740, 780, or 800 (SEQ ID NO:2), resulting from deletions that eliminate 237, 217, 197, 117, 77 and 57 amino acids, respectively, relative to the full-length β RT. Construction and expression of a deletion derivative specifying a modified β-like RT of 620 amino acids was accomplished as generally described in Examples 3 and 4, with the expression results presented below.

M1BA 2000-3000 U/g

M1KA 1000 U/L

M1BK 620 31,950 U/g

Thus, a truncated β-like RT shows considerable activity, consistent with an increase in solubility relative to the full-length native β RT.

Analogous modifications to the corresponding β polypeptides of other avian RTs result in similarly increased RT activity.

RSV-RT 620 his 33,000 U/g

In addition to 3' deletions resulting in polynucleotides encoding β-like polypeptides having C-termini in the range of positions 580-832, and preferably in the range of 620-800 (SEQ ID NO:2), the invention contemplates polynucleotides having internal deletions relative to the native β gene, as well as the polypeptides encoded by polynucleotides having such internal deletions. The central core region of the integrase domain is associated with the DNA cutting and joining properties of the native AMV-RT.

The core region of the integrase domain was deleted to varying extents (the region between amino acids 620-770, 640-770 or 660-770 of SEQ ID NO:2), *e.g.*, M1BK Cint lacks amino acids 620-770 of SEQ ID NO:2, using conventional techniques. The approach involved the initial construction of first polynucleotide fragments encoding C-terminally truncated β-like fragments using PCR with the full-length AMV-RT *pol* gene as a template (see Table II). Second fragments containing various lengths the 3' of the end of the *pol* gene

(*i.e.*, 3' fragment) were also constructed using PCR. These 3' fragments encoded the C-terminal region of the integrase domain, some 3' fragments also contained part, but not all, of the core region of the integrase domain. Those of skill in the art will recognize that the first polynucleotide fragments, or 5' fragments, may encode peptide tags at their 5' ends; the 3' fragments may also encode peptide tags (see *e.g.*, 3' Fragment 2a in Table II), with or without tags encoded by the 5' fragment, and these tag-encoding fragments are readily synthesized using the PCR primers disclosed herein (*e.g.*, F Cint XhoI (SEQ ID NO:85) and R Cint 830 His XhoI (SEQ ID NO:91)). The final step in generating constructs having internal deletions was to ligate truncated β-like coding regions to 3' fragments in proper order and orientation, as determined by the conventional screening of ligation products. In one embodiment, amino acids 620-770 were deleted, thereby removing the core region of the integrase domain. The C-terminal region of the integrase domain was then placed adjacent to the N-terminal region of that domain.

Expression of such constructs in insect cells revealed an increase in solubility (10-20%) and activity relative to the full-length, intact β RT, as shown below. Other deletions effectively removing part or all of the central region of the integrase domain, such as the deletion of amino acids 620-731, 640-771, 640-731, 660-771, 660-731, 680-771, 680-731, and 740-771 (SEQ ID NO:2) are contemplated by the invention.

M1KA	1000 U/L
M1-5	200 U/L

Some modified beta fragments have terminal peptide tags. Thus, the invention contemplates modified RTs having internal deletions and, optionally, peptide tags at an N-terminus, a C-terminus or both termini. In addition, as for α-like modified RTs, the β-like modified RTs may be derived from any Type II or Type III RT, along with polynucleotides encoding them.

Any of the modified RTs of the invention may be produced by any process disclosed herein or known in the art, such as *in vivo* synthesis, *in vitro* synthesis or chemical synthesis. Further, any of these processes may be used to produce active polypeptides in a variety of forms, including monomers, homo-dimers or homo-multimers, hetero-dimers, and hetero-multimers, all of which are comprehended by the invention. In particular, expression of the modified beta-like fragment M1BK620 Cint resulted in expression of a heterodimeric form of RT, suggesting that the beta-like fragment was processed as expected, to yield an α polypeptide in association with a modified β-like polypeptide. Expression of other modified

RTs of the invention, such as other core domain deletions (*e.g.*,β-like fragments lacking amino acids 620-731, 640-771, 640-731, 660-771, 660-731, 680-771, 680-731, or 740-771 of SEQ ID NO 2) are expected to show activity in other than monomeric form, *e.g.*, in heterodimeric form. In addition, heterodimers or other non-monomeric forms may arise from the interaction of a modified α-like polypeptide and a native β polypeptide, or from a modified α-like polypeptide and a modified β-like polypeptide, regardless of whether the polypeptides were produced by *in vivo* or *in vitro* expression, or by chemical synthesis.

While the present invention has been described in terms of specific embodiments, it is understood that variations and modifications will occur to those skilled in the art. Accordingly, only those limitations appearing in the appended claims should be placed upon the invention.

What is claimed is:

1 An isolated polynucleotide encoding a polypeptide having RNA-dependent DNA polymerase activity, the polypeptide consisting of

(a) an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 428 to 857 of SEQ ID NO:2;

(b) an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 428 to 1054 of SEQ ID NO:39;

(c) an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 548 to 1198 of SEQ ID NO:41;

(d) an amino acid sequence beginning at amino acid 1 and terminating at any one of amino acids 428 to 901 of SEQ ID NO:43; and

(e) variants, analogs and fragments of any of subparts (a) to (e) having RNA-dependent DNA polymerase activity,

said polypeptide, variants, analogs, and fragments optionally having an N-terminal methionine.

2. The polynucleotide according to claim 1 step (a) wherein said polypeptide consists of a sequence that begins at about amino acid 1 and ends at about amino acid 578 of SEQ ID NO:2.

3. The polynucleotide according to claim 1 step (a) wherein said polypeptide consists of the sequence set forth as SEQ ID NO:4.

4. The polynucleotide according to claim 1 having a sequence selected from the group consisting of a sequence set forth in any one of SEQ ID NOs 1, 6-10, 38, 40, and 42.

5. The polynucleotide according to claim 1 wherein said polynucleotide is DNA.

6. The polynucleotide according to claim 1 wherein said polynucleotide encodes a polypeptide that lacks an effective integrase activity.

7. The polynucleotide according to claim 6 wherein said polynucleotide lacks at least part of an integrase coding region.

8. The polynucleotide according to claim 1 further comprising an adjacent polynucleotide encoding at least one terminal modification of said polypeptide selected from the group consisting of an N-terminal modification and a C-terminal modification.

9. The polynucleotide according to claim 8 wherein said modification is a cysteine residue adjacent the C-terminus of said polypeptide.

10. The polynucleotide according to claim 8 wherein said adjacent polynucleotide encodes a polypeptide consisting of a C-terminal modification.

11. The polynucleotide according to claim 10 wherein said C-terminal polypeptide comprises between four and fifty amino acids and wherein said polypeptide comprises a domain selected from the group consisting of a DNA binding domain, an RNA binding domain, a metal binding domain, a structure stabilizing domain, and a polymerizing domain.

12. The polynucleotide according to claim 11 wherein said polypeptide comprises an acidic amino acid domain, a basic amino acid domain, a W domain, a WH domain, a zinc-finger-like domain, a leucine zipper domain, a PPG domain, an NS1 domain, a GPRP domain, and a PRPG domain.

13. The polynucleotide according to claim 11 wherein said C-terminal peptide comprises six amino acids.

14. The polynucleotide according to claim 11 wherein said C-terminal peptide comprises amino acids that are the same.

15. The polynucleotide according to claim 11 wherein said C-terminal peptide comprises amino acids that are basic.

16. The polynucleotide according to claim 15 wherein said basic amino acids are histidine.

17. The polynucleotide according to claim 8 having a sequence selected from the group consisting of a sequence set forth in any one of SEQ ID NOs 11-19.

18. A vector comprising the polynucleotide according to claim 1.

19. The vector according to claim 18 wherein said polynucleotide is operably linked to a promoter.

20. A host cell transformed with a vector according to claim 18.

21. The host cell according to claim 20 wherein said host cell is a eukaryotic cell.

22. The host cell according to claim 20 wherein said host cell is selected from the group consisting of *Escherichia coli* and an insect cell.

23. An isolated polypeptide encoded by the polynucleotide according to any one of claims 1 to 5.

24. An isolated polypeptide encoded by the polynucleotide according to any one of claims 6 to 17.

25. A method of transforming host cells comprising the following steps:

(a) introducing a vector according to claim 18 into host cells;

(b) incubating said host cells; and

(c) identifying host cells containing said vector, thereby identifying a transformed host cell.

26. A method of producing an isolated Reverse Transcriptase polypeptide comprising the following steps:

(a) transforming a host cell with a vector according to claim 18;

(b) incubating said host cell under conditions suitable for expression of a polypeptide; and

(c) recovering said polypeptide, thereby producing an isolated Reverse Transcriptase.

27. In a method for copying a target nucleic acid by extending a target nucleic acid-bound primer in the presence of a polymerase, the improvement comprising:

(a) contacting said target nucleic acid and primer with the polypeptide according to any one of claims 23 and 24.

28. The method according to claim 27 wherein said copying produces a plurality of copies of said target nucleic acid.

29. The method according to claim 27 wherein said polypeptide is in a form selected from the group consisting of a monomer and a polymer.

30. The method according to claim 27 wherein said method is selected from the group consisting of cDNA synthesis, Polymerase Chain Reaction, Polymerase Chain Reaction-Reverse Transcription, Inverse Polymerase Chain Reaction, Multiplex Polymerase Chain Reaction, Strand Displacement Amplification, Multiplex Strand Displacement Amplification, Nucleic Acid Sequence-Based Amplification, Sequence-Specific Strand Replication and Rapid Amplification.

31. In a method for sequencing a target nucleic acid by extending a target nucleic acid-bound primer, the improvement comprising:

(a) contacting said target nucleic acid and primer with the polypeptide according to any one of claims 23 and 24.

32. The method according to claim 31 wherein said polypeptide is in a form selected from the group consisting of a monomer and a polymer.

33. A kit for copying a target nucleic acid comprising:

(a) one or more nucleotides, and

(b) a polypeptide encoded by a polynucleotide having a sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO 8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO 13, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42 and polynucleotide derivatives thereof encoding C-terminal modifications at their 3′ ends.

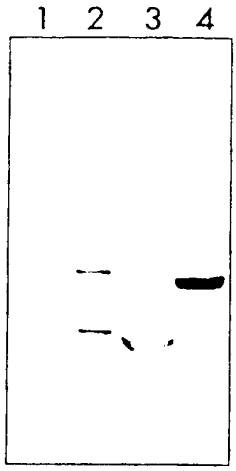

Fig. 1

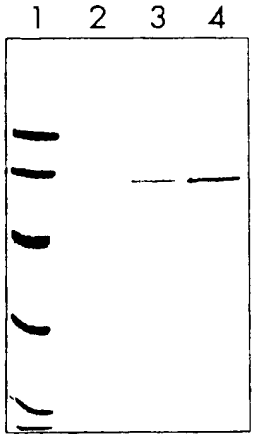

Fig. 2

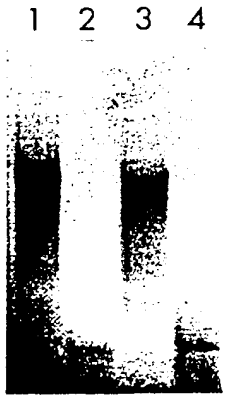

Fig. 3

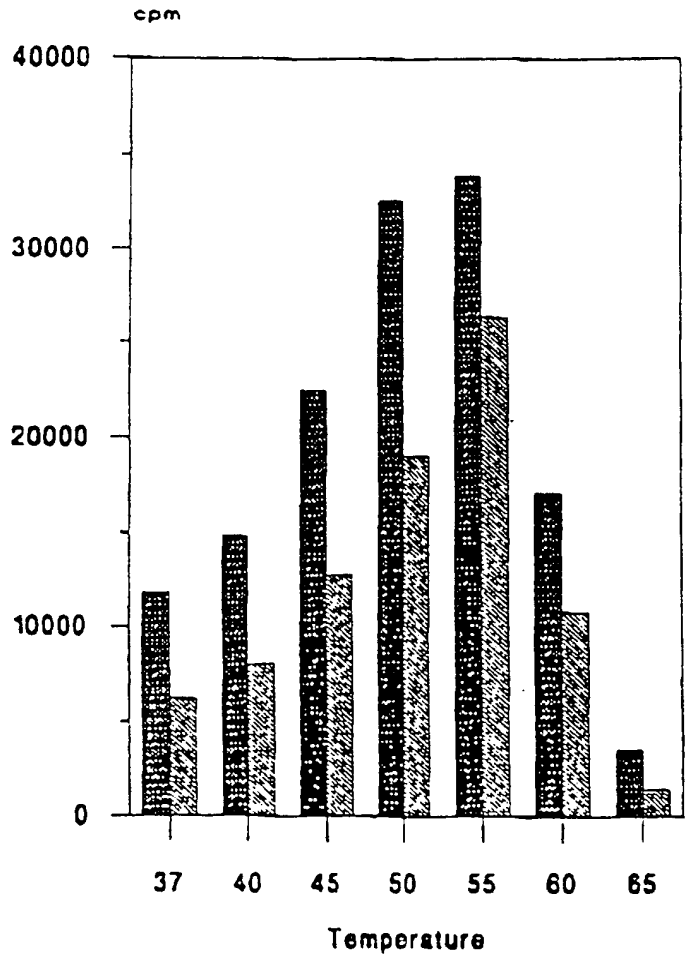

Fig. 4A

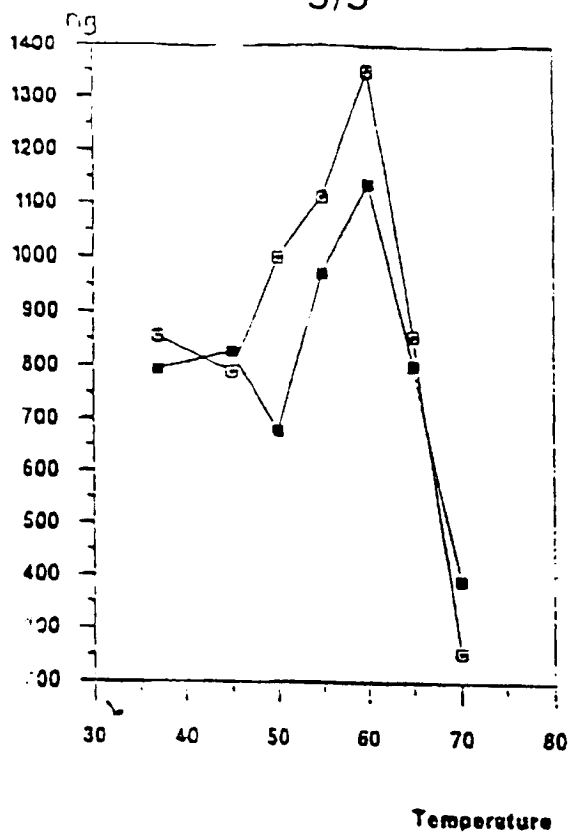

Fig. 4B

RAMP Temperature Profile

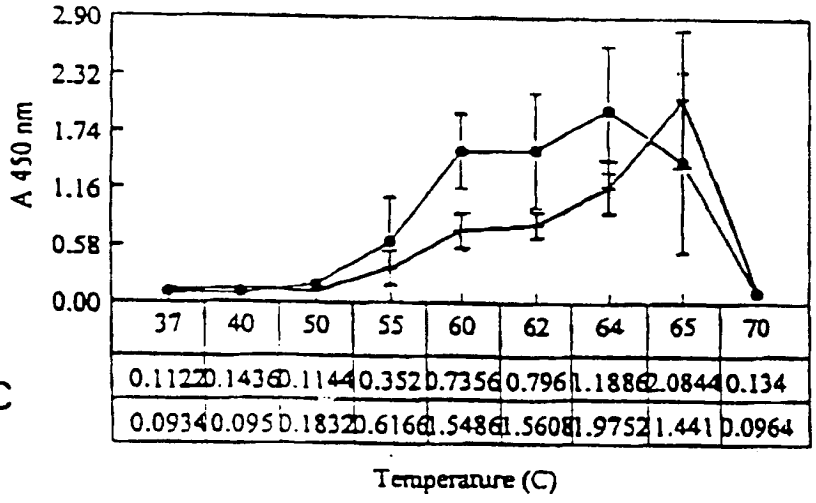

Fig. 4C

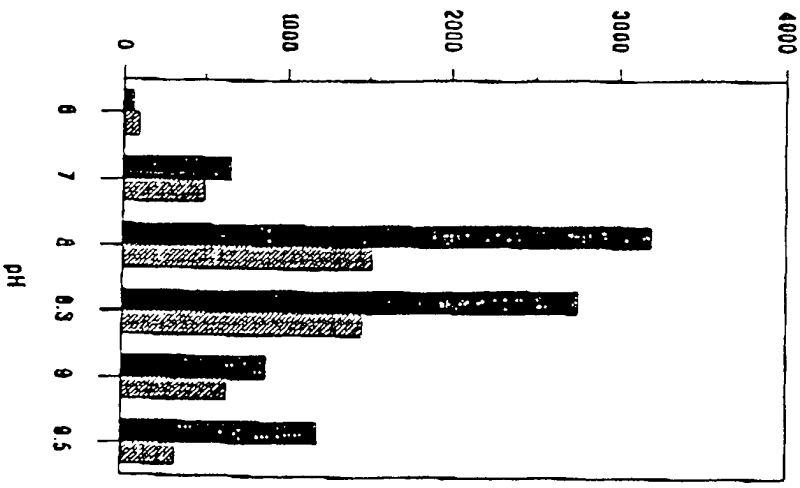
RT Assay at Various Tris pH (60)
cpm
0
1000
2000
3000
4000
pH
6
7
8
8.3
9
9.5

Fig. 4D

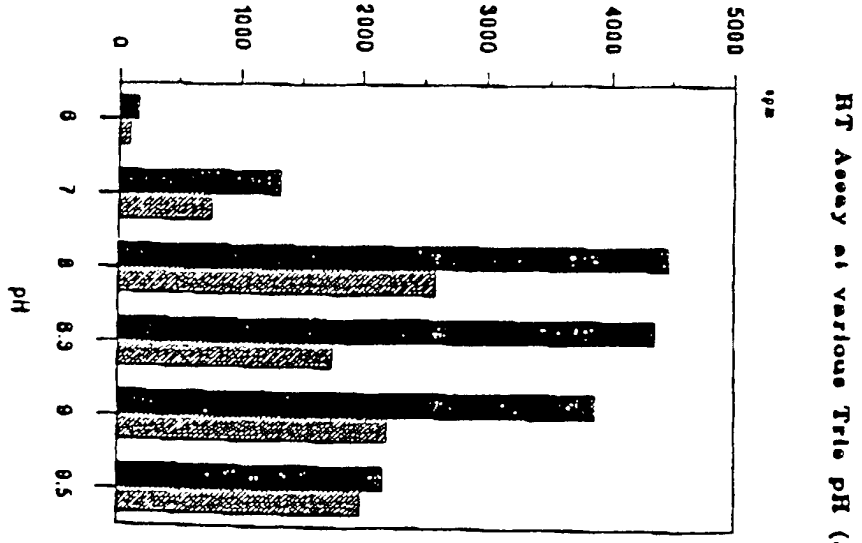
RT Assay at various Tris pH (37)
cpm
0
1000
2000
3000
4000
5000
pH
6
7
8
8.3
9
9.5

Fig. 4E

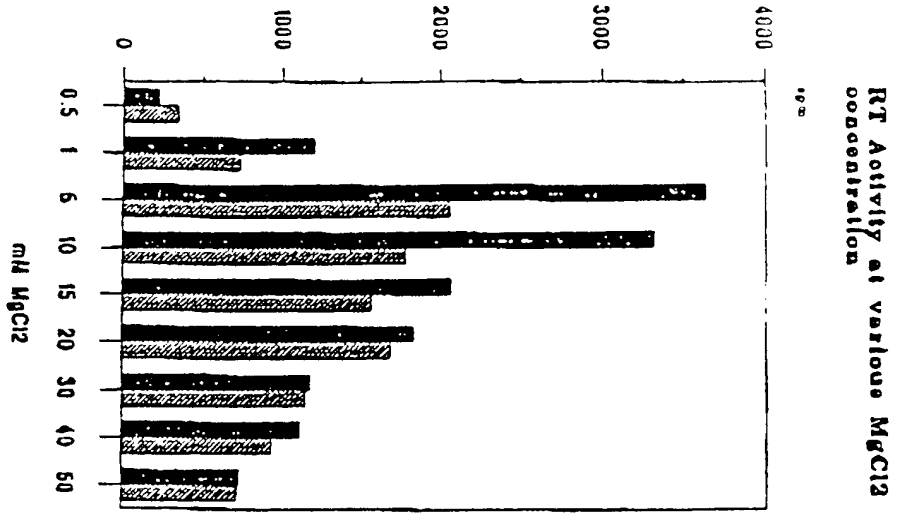
RT Activity at various MgCl2 concentration
0
1000
2000
3000
4000
0.5
1
5
10
15
20
30
40
50
mM MgCl2

Fig. 4F

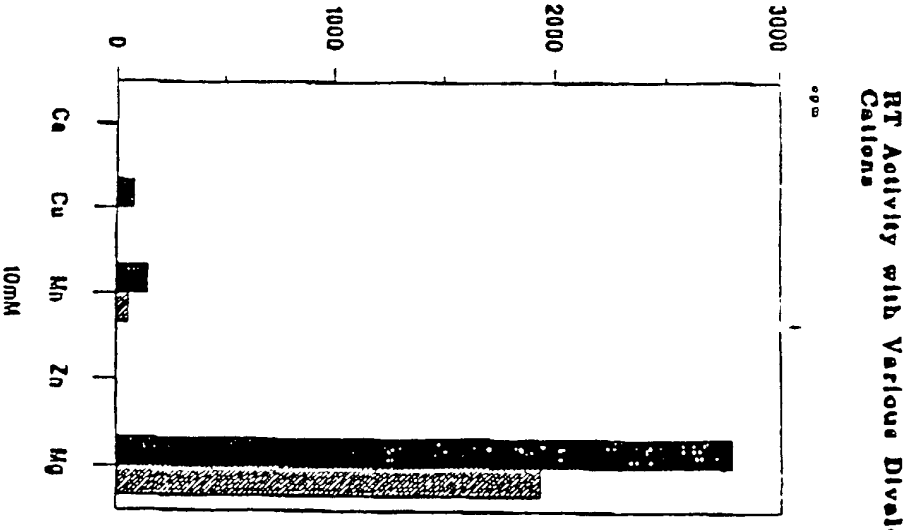
RT Activity with Various Divalent Cations
0
1000
2000
3000
Ca
Cu
Mn
Zn
Mg
10mM

Fig. 4G

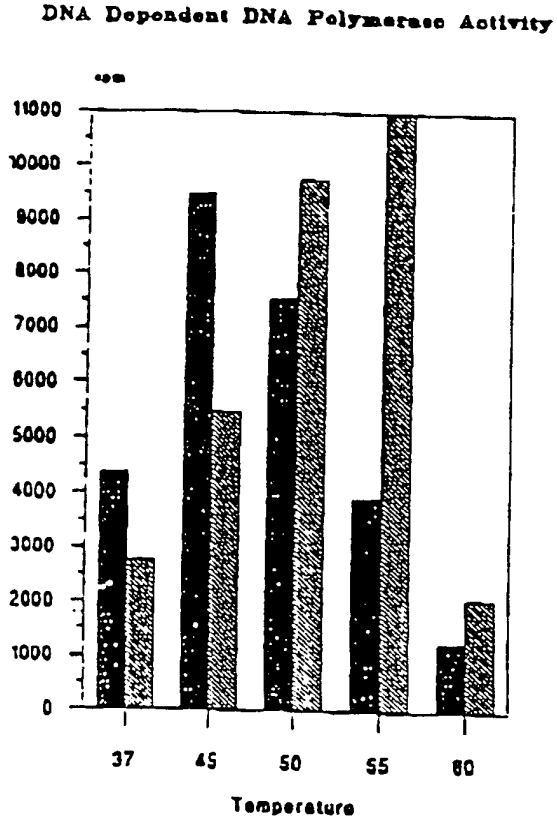
DNA Dependent DNA Polymerase Activity
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
37
45
50
55
60
Temperature

Fig. 5

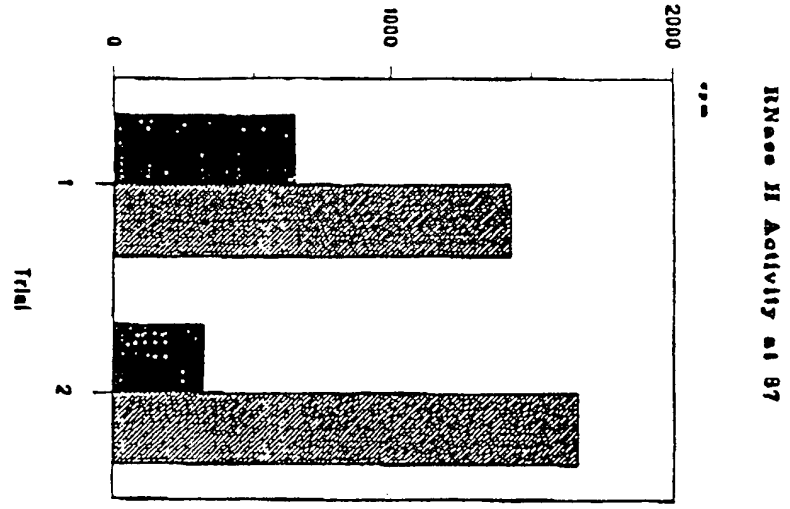
RNase H Activity at 37
0
1000
2000
Trial
1
2

Fig. 6A

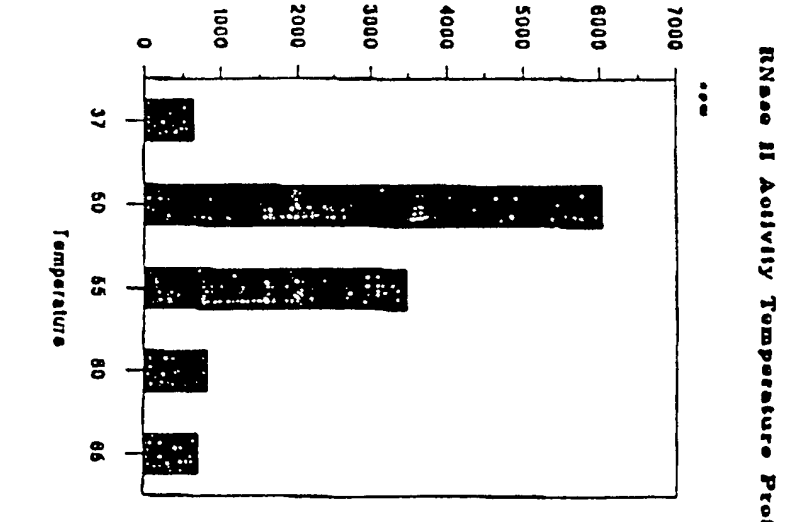
RNase H Activity Temperature Profile
0
1000
2000
3000
4000
5000
6000
7000
Temperature
37
50
55
80
95

Fig. 6B

SEQUENCE LISTING

<110> Swaminathan, Nanda (inventor)
MOLECULAR BIOLOGY RESOURCES, INC.

<120> BIOLOGICALLY ACTIVE REVERSE TRANSCRIPTASES

<130> 28003/34938

<140>
<141>

<150> 60/116,099
<151> 1999-01-15

<160> 100

<170> PatentIn Ver. 2.0

<210> 1
<211> 3152
<212> DNA
<213> myeloblastosis-associated virus

<220>
<221> CDS
<222> (253)..(2937)
<223> full-length MAV-RT translated polynucleotide

<400> 1
ggaggattgg cccaccgatt ggccagtgat ggaggccgcg aacccgcaga tccatgggat 60

aggaggggga attcccatgc gaaaatctcg tgacatgata gagttggggg ttattaaccg 120

agacgggtct ttggagcgac ccctgctcct cttccccgca gtagctatgg ttagagggag 180

tatcctagga agagattgtc tgcagggcct agggctccgc ttgacaaatt tatagggagg 240

gccactgttc tc act gtt gcg cta cat ctg gct att ccg ctc aaa tgg aag 291
Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys
1 5 10

```
cca aac cat aca cct gtg tgg att gac cag tgg ccc ctt cct gaa ggt   339
Pro Asn His Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly
 15                  20                  25

aaa ctt gta gcg cta acg caa tta gtg gaa aaa gaa tta cag tta gga   387
Lys Leu Val Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly
 30                  35                  40                  45

cat ata gaa cct tca ctt agt tgc tgg aac aca cct gtc ttt gtg atc   435
His Ile Glu Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile
                 50                  55                  60

cgg aag gct tca ggt tct tat cgc tta ttg cat gac ttg cgc gct gtt   483
Arg Lys Ala Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val
             65                  70                  75

aac gct aag ctt gtt cct ttt ggg gcc gtc caa cag ggg gcg ccg gtt   531
Asn Ala Lys Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val
         80                  85                  90

ctc tcc gcg ctc ccg cgt ggt tgg ccc ctg atg gtc cta gac ctc aag   579
Leu Ser Ala Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys
 95                  100                 105

gat tgc ttc ttt tct att cct ctt gcg gaa caa gat cgc gaa cgt ttt   627
Asp Cys Phe Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Arg Phe
110                 115                 120                 125

gca ttt acg ctc ccc tcc gtg aat aac cag gcc ccc gct cga agg ttc   675
Ala Phe Thr Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe
                130                 135                 140

caa tgg aag gtc ttg ccc caa ggg atg acc tgt tct ccc act atc tgt   723
Gln Trp Lys Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys
            145                 150                 155

cag ttg ata gtg ggt caa ata ctt gag ccc ttg cga ctc aag cac cca   771
Gln Leu Ile Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro
        160                 165                 170

tct ctg cgc atg ttg cat tat atg gat gat ctt ttg cta gcc gcc tca   819
Ser Leu Arg Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser
    175                 180                 185
```

agt cat gat ggg ttg gaa gcg gca ggg gag gag gtt atc agt aca ttg 867
Ser His Asp Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu
190 195 200 205

gaa aga gcc ggg ttc acc att tcg cct gat aag gtc cag agg gag ccc 915
Glu Arg Ala Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Arg Glu Pro
210 215 220

gga gta caa tat ctt ggg tac aag tta ggt agt acg tat gta gca ccc 963
Gly Val Gln Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Val Ala Pro
225 230 235

gta gtc ctg gta gca gaa ccc agg ata gcc acc ttg tgg gat gtt cag 1011
Val Gly Leu Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln
240 245 250

aag ctg gtg ggg tca ctt cag tgg ctt cgc cca gcg tta gga atc ccg 1059
Lys Leu Val Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro
255 260 265

cca cga ctg atg ggc ccc ttt tat gag cag tta cga ggg tca gat cct 1107
Pro Arg Leu Met Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro
270 275 280 285

aac gag gcg agg gaa tgg aat cta gac atg aaa atg gcc tgg aga gag 1155
Asn Glu Ala Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu
290 295 300

atc gta cag ctc agc acc act gct gcc ttg gag cga tgg gac cct gcc 1203
Ile Val Gln Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala
305 310 315

ctg cct ctg gaa gga gcg gtc gct aga tgt gaa cag ggg gca ata ggg 1251
Leu Pro Leu Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly
320 325 330

gtc ctg gga cag gga ctg tcc aca cac cca agg cca tgt ttg tgg cta 1299
Val Leu Gly Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu
335 340 345

ttc tcc acc caa ccc acc aag gcg ttt act gct tgg tta gaa gtg ctc 1347
Phe Ser Thr Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu
350 355 360 365

```
acc ctt ttg att act aag cta cgt gct tcg gca gtg cga acc ttt ggc  1395
Thr Leu Leu Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly
                370                 375                 380

aag gag gtt gat atc ctc ctg ttg cct gca tgc ttt cgg gag gac ctt  1443
Lys Glu Val Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Glu Asp Leu
            385                 390                 395

ccg ctc ccg gag ggg atc ctg tta gcc ctt aag ggg ttt gca gga aaa  1491
Pro Leu Pro Glu Gly Ile Leu Leu Ala Leu Lys Gly Phe Ala Gly Lys
        400                 405                 410

atc agg agt agt gac acg cca tct att ttt gac att gcg cgt cca ctg  1539
Ile Arg Ser Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu
    415                 420                 425

cat gtt tct ctg aaa gtg agg gtt acc gac cac cct gtg ccg gga ccc  1587
His Val Ser Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro
430                 435                 440                 445

act gtc ttt act gac gcc tcc tca agc acc cat aag ggg gtg gta gtc  1635
Thr Val Phe Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val
                450                 455                 460

tgg agg gag ggc cca agg tgg gag ata aaa gaa ata gct gat ttg ggg  1683
Trp Arg Glu Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly
            465                 470                 475

gca agt gta caa caa ctg gaa gca cgc gct gtg gcc atg gca ctt ctg  1731
Ala Ser Val Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu
        480                 485                 490

ctg tgg ccg aca acg ccc act aat gta gtg act gac tcc gcg ttt gtt  1779
Leu Trp Pro Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val
    495                 500                 505

gcg aaa atg tta ctc aag atg gga cag gag gga gtc ccg tct aca gcg  1827
Ala Lys Met Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala
510                 515                 520                 525

gcg gct ttt att tta gag gat gcg tta agc caa agg tca gcc atg gcc  1875
Ala Ala Phe Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala
                530                 535                 540
```

gcc gtt ctc cac gtg cgg agt cat tct gaa gtg cca ggg ttt ttc aca 1923
Ala Val Leu His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr
545 550 555

gaa gga aat gac gtg gca gat agc caa gcc acc ttt caa gcg tat ccc 1971
Glu Gly Asn Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro
560 565 570

ttg aga gag gct aaa gat ctc cat acc gct ctc cat att gga ccc cgc 2019
Leu Arg Glu Ala Lys Asp Leu His Thr Ala Leu His Ile Gly Pro Arg
575 580 585

gcg cta tcc aaa gcg tgt aat ata tct atg cag cag gct agg gag gtt 2067
Ala Leu Ser Lys Ala Cys Asn Ile Ser Met Gln Gln Ala Arg Glu Val
590 595 600 605

gtt cag acc tgc ccg cat tgt aat tca gcc cct gcg ttg gag gcc ggg 2115
Val Gln Thr Cys Pro His Cys Asn Ser Ala Pro Ala Leu Glu Ala Gly
610 615 620

gta aac cct agg ggt ttg gga ccc cta cag ata tgg cag aca gac ttt 2163
Val Asn Pro Arg Gly Leu Gly Pro Leu Gln Ile Trp Gln Thr Asp Phe
625 630 635

aca ctt gag cct aga atg gcc ccc cgt tcc tgg ctc gct gtt act gtg 2211
Thr Leu Glu Pro Arg Met Ala Pro Arg Ser Trp Leu Ala Val Thr Val
640 645 650

gat acc gcc tca tcg gcg ata gtc gta act cag cat ggc cgt gtc aca 2259
Asp Thr Ala Ser Ser Ala Ile Val Val Thr Gln His Gly Arg Val Thr
655 660 665

tcg gtt gct gca caa cat cat tgg gcc acg gct atc gcc gtt ttg gga 2307
Ser Val Ala Ala Gln His His Trp Ala Thr Ala Ile Ala Val Leu Gly
670 675 680 685

aga cca aag gcc ata aaa aca gat aat ggg tcc tgc ttc acg tct aaa 2355
Arg Pro Lys Ala Ile Lys Thr Asp Asn Gly Ser Cys Phe Thr Ser Lys
690 695 700

tcc acg cga gag tgg ctc gcg aga tgg ggg ata gca cac acc acc ggg 2403
Ser Thr Arg Glu Trp Leu Ala Arg Trp Gly Ile Ala His Thr Thr Gly
705 710 715

att ccg ggt aat tcc cag ggt caa gct atg gta gag cgg gcc aac cgg 2451
Ile Pro Gly Asn Ser Gln Gly Gln Ala Met Val Glu Arg Ala Asn Arg
720 725 730

ctc ctg aaa gat aag atc cgt gtg ctt gcg gag ggg gat ggc ttt atg 2499
Leu Leu Lys Asp Lys Ile Arg Val Leu Ala Glu Gly Asp Gly Phe Met
735 740 745

aaa aga atc ccc acc agc aaa cag ggg gaa cta tta gcc aag gca atg 2547
Lys Arg Ile Pro Thr Ser Lys Gln Gly Glu Leu Leu Ala Lys Ala Met
750 755 760 765

tat gcc ctc aat cac ttt gag cgt ggt gaa aac aca aaa aca ccg ata 2595
Tyr Ala Leu Asn His Phe Glu Arg Gly Glu Asn Thr Lys Thr Pro Ile
770 775 780

caa aaa cac tgg aga cct acc gtt ctt aca gaa gga ccc ccg gtt aaa 2643
Gln Lys His Trp Arg Pro Thr Val Leu Thr Glu Gly Pro Pro Val Lys
785 790 795

ata cga ata gag aca ggg gag tgg gaa aaa gga tgg aac gtg ctg gtc 2691
Ile Arg Ile Glu Thr Gly Glu Trp Glu Lys Gly Trp Asn Val Leu Val
800 805 810

tgg gga cga ggt tat gcc gct gtg aaa aac agg gac act gat aag gtt 2739
Trp Gly Arg Gly Tyr Ala Ala Val Lys Asn Arg Asp Thr Asp Lys Val
815 820 825

att tgg gta ccc tct cga aaa gtt aaa ccg gac atc acc caa aag gat 2787
Ile Trp Val Pro Ser Arg Lys Val Lys Pro Asp Ile Thr Gln Lys Asp
830 835 840 845

gag gtg act aag aaa gat gag gcg agc cct ctt ttt gca ggc att tct 2835
Glu Val Thr Lys Lys Asp Glu Ala Ser Pro Leu Phe Ala Gly Ile Ser
850 855 860

gac tgg gcg ccc tgg gaa ggc gag caa gaa gga ctc caa gaa gaa acc 2883
Asp Trp Ala Pro Trp Glu Gly Glu Gln Glu Gly Leu Gln Glu Glu Thr
865 870 875

gcc agc aac aag caa gaa aga ccc gga gaa gac acc cct gct gcc aac 2931
Ala Ser Asn Lys Gln Glu Arg Pro Gly Glu Asp Thr Pro Ala Ala Asn
880 885 890

gag agt taattatatt ctcattattg gtgtcctggt cttgtgtgag gttacggggg 2987
Glu Ser
895

taagagcctg aatgttcact tactcgagca gccagggaac ctttggatta catgggccaa 3047

ccgtacaggc caaacggatt tctgcctctc tacacagtca gccacctccc cttttcaaac 3107

atgtttgata ggtatcccgt ctcctatttc cgaaggtgat tttaa 3152

<210> 2
<211> 895
<212> PRT
<213> myeloblastosis-associated virus

<400> 2

Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys Pro Asn His
1 5 10 15

Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly Lys Leu Val
20 25 30

Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly His Ile Glu
35 40 45

Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile Arg Lys Ala
50 55 60

Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val Asn Ala Lys
65 70 75 80

Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val Leu Ser Ala
85 90 95

Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys Asp Cys Phe
100 105 110

Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Arg Phe Ala Phe Thr
115 120 125

Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe Gln Trp Lys
130 135 140

Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys Gln Leu Ile
145 150 155 160

Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro Ser Leu Arg
165 170 175

Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser Ser His Asp
180 185 190

Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu Glu Arg Ala
195 200 205

Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Arg Glu Pro Gly Val Gln
210 215 220

Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Val Ala Pro Val Gly Leu
225 230 235 240

Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln Lys Leu Val
245 250 255

Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro Pro Arg Leu
260 265 270

Met Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro Asn Glu Ala
275 280 285

Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu Ile Val Gln
290 295 300

Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala Leu Pro Leu
305 310 315 320

Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly Val Leu Gly
325 330 335

Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu Phe Ser Thr
340 345 350

Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu Thr Leu Leu
355 360 365

Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly Lys Glu Val
370 375 380

Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Glu Asp Leu Pro Leu Pro
385 390 395 400

Glu Gly Ile Leu Leu Ala Leu Lys Gly Phe Ala Gly Lys Ile Arg Ser
405 410 415

Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu His Val Ser
420 425 430

Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro Thr Val Phe
435 440 445

Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val Trp Arg Glu
450 455 460

Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly Ala Ser Val
465 470 475 480

Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu Leu Trp Pro
485 490 495

Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val Ala Lys Met
500 505 510

Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala Ala Ala Phe
515 520 525

Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala Ala Val Leu
530 535 540

His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr Glu Gly Asn
545 550 555 560

Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro Leu Arg Glu
565 570 575

Ala Lys Asp Leu His Thr Ala Leu His Ile Gly Pro Arg Ala Leu Ser
580 585 590

Lys Ala Cys Asn Ile Ser Met Gln Gln Ala Arg Glu Val Val Gln Thr
595 600 605

Cys Pro His Cys Asn Ser Ala Pro Ala Leu Glu Ala Gly Val Asn Pro
610 615 620

Arg Gly Leu Gly Pro Leu Gln Ile Trp Gln Thr Asp Phe Thr Leu Glu
625 630 635 640

Pro Arg Met Ala Pro Arg Ser Trp Leu Ala Val Thr Val Asp Thr Ala
645 650 655

Ser Ser Ala Ile Val Val Thr Gln His Gly Arg Val Thr Ser Val Ala
660 665 670

Ala Gln His His Trp Ala Thr Ala Ile Ala Val Leu Gly Arg Pro Lys
675 680 685

Ala Ile Lys Thr Asp Asn Gly Ser Cys Phe Thr Ser Lys Ser Thr Arg
690 695 700

Glu Trp Leu Ala Arg Trp Gly Ile Ala His Thr Thr Gly Ile Pro Gly
705 710 715 720

Asn Ser Gln Gly Gln Ala Met Val Glu Arg Ala Asn Arg Leu Leu Lys
725 730 735

Asp Lys Ile Arg Val Leu Ala Glu Gly Asp Gly Phe Met Lys Arg Ile
740 745 750

Pro Thr Ser Lys Gln Gly Glu Leu Leu Ala Lys Ala Met Tyr Ala Leu
755 760 765

Asn His Phe Glu Arg Gly Glu Asn Thr Lys Thr Pro Ile Gln Lys His
770 775 780

Trp Arg Pro Thr Val Leu Thr Glu Gly Pro Pro Val Lys Ile Arg Ile
785 790 795 800

Glu Thr Gly Glu Trp Glu Lys Gly Trp Asn Val Leu Val Trp Gly Arg
805 810 815

Gly Tyr Ala Ala Val Lys Asn Arg Asp Thr Asp Lys Val Ile Trp Val
820 825 830

Pro Ser Arg Lys Val Lys Pro Asp Ile Thr Gln Lys Asp Glu Val Thr
835 840 845

Lys Lys Asp Glu Ala Ser Pro Leu Phe Ala Gly Ile Ser Asp Trp Ala
850 855 860

Pro Trp [illegible] Gly Glu Gln Glu Gly Leu Gln Glu Glu Thr Ala Ser Asn
865 870 875 880

Lys Gln [illegible] Arg [illegible] Gly Glu Asp Thr Pro Ala Ala Asn Glu Ser
885 890 895

<210> 3
<211> 896
<212> PRT
<213> myeloblastosis-associated virus

<220>
<223> full length + Met- amino acid

<400> 3
Met Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys Pro Asn
1 5 10 15

His Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly Lys Leu
20 25 30

Val Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly His Ile
35 40 45

Glu Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile Arg Lys
50 55 60

Ala Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val Asn Ala
65 70 75 80

Lys Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val Leu Ser
85 90 95

Ala Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys Asp Cys
100 105 110

Phe Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Arg Phe Ala Phe
115 120 125

Thr Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe Gln Trp
130 135 140

Lys Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys Gln Leu

145 150 155 160

Ile Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro Ser Leu
165 170 175

Arg Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser Ser His
180 185 190

Asp Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu Glu Arg
195 200 205

Ala Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Arg Glu Pro Gly Val
210 215 220

Gln Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Val Ala Pro Val Gly
225 230 235 240

Leu Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln Lys Leu
245 250 255

Val Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro Pro Arg
260 265 270

Leu Met Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro Asn Glu
275 280 285

Ala Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu Ile Val
290 295 300

Gln Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala Leu Pro
305 310 315 320

Leu Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly Val Leu
325 330 335

Gly Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu Phe Ser
340 345 350

Thr Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu Thr Leu
355 360 365

Leu Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly Lys Glu
370 375 380

Val Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Glu Asp Leu Pro Leu
385 390 395 400

Pro Glu Gly Ile Leu Leu Ala Leu Lys Gly Phe Ala Gly Lys Ile Arg
405 410 415

Ser Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu His Val
420 425 430

Ser Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro Thr Val
435 440 445

Phe Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val Trp Arg
450 455 460

Glu Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly Ala Ser
465 470 475 480

Val Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu Leu Trp
485 490 495

Pro Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val Ala Lys
500 505 510

Met Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala Ala Ala
515 520 525

Phe Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala Ala Val
530 535 540

Leu His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr Glu Gly
545 550 555 560

Asn Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro Leu Arg
565 570 575

Glu Ala Lys Asp Leu His Thr Ala Leu His Ile Gly Pro Arg Ala Leu
580 585 590

Ser Lys Ala Cys Asn Ile Ser Met Gln Gln Ala Arg Glu Val Val Gln
595 600 605

Thr Cys Pro His Cys Asn Ser Ala Pro Ala Leu Glu Ala Gly Val Asn
610 615 620

Pro Arg Gly Leu Gly Pro Leu Gln Ile Trp Gln Thr Asp Phe Thr Leu
625 630 635 640

Glu Pro Arg Met Ala Pro Arg Ser Trp Leu Ala Val Thr Val Asp Thr
645 650 655

Ala Ser Ser Ala Ile Val Val Thr Gln His Gly Arg Val Thr Ser Val
660 665 670

Ala Ala Gln His His Trp Ala Thr Ala Ile Ala Val Leu Gly Arg Pro
675 680 685

Lys Ala Ile Lys Thr Asp Asn Gly Ser Cys Phe Thr Ser Lys Ser Thr
690 695 700

Arg Glu Trp Leu Ala Arg Trp Gly Ile Ala His Thr Thr Gly Ile Pro
705 710 715 720

Gly Asn Ser Gln Gly Gln Ala Met Val Glu Arg Ala Asn Arg Leu Leu
725 730 735

Lys Asp Lys Ile Arg Val Leu Ala Glu Gly Asp Gly Phe Met Lys Arg
740 745 750

Ile Pro Thr Ser Lys Gln Gly Glu Leu Leu Ala Lys Ala Met Tyr Ala
755 760 765

Leu Asn His Phe Glu Arg Gly Glu Asn Thr Lys Thr Pro Ile Gln Lys
770 775 780

His Trp Arg Pro Thr Val Leu Thr Glu Gly Pro Pro Val Lys Ile Arg
785 790 795 800

Ile Glu Thr Gly Glu Trp Glu Lys Gly Trp Asn Val Leu Val Trp Gly
805 810 815

Arg Gly Tyr Ala Ala Val Lys Asn Arg Asp Thr Asp Lys Val Ile Trp
820 825 830

Val Pro Ser Arg Lys Val Lys Pro Asp Ile Thr Gln Lys Asp Glu Val
835 840 845

Thr Lys Lys Asp Glu Ala Ser Pro Leu Phe Ala Gly Ile Ser Asp Trp
850 855 860

Ala Pro Trp Glu Gly Glu Gln Glu Gly Leu Gln Glu Glu Thr Ala Ser
865 870 875 880

Asn Lys Gln Glu Arg Pro Gly Glu Asp Thr Pro Ala Ala Asn Glu Ser
885 890 895

<210> 4
<211> 578
<212> PRT
<213> myeloblastosis-associated virus

<220>
<223> alpha (no met, no tag, no stop)

<400> 4
Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys Pro Asn His
1 5 10 15

Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly Lys Leu Val
20 25 30

Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly His Ile Glu
35 40 45

Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile Arg Lys Ala
50 55 60

Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val Asn Ala Lys
65 70 75 80

Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val Leu Ser Ala
85 90 95

Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys Asp Cys Phe
100 105 110

Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Arg Phe Ala Phe Thr
115 120 125

Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe Gln Trp Lys

130 135 140

Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys Gln Leu Ile
145 150 155 160

Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro Ser Leu Arg
165 170 175

Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser Ser His Asp
180 185 190

Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu Glu Arg Ala
195 200 205

Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Arg Glu Pro Gly Val Gln
210 215 220

Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Val Ala Pro Val Gly Leu
225 230 235 240

Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln Lys Leu Val
245 250 255

Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro Pro Arg Leu
260 265 270

Met Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro Asn Glu Ala
275 280 285

Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu Ile Val Gln
290 295 300

Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala Leu Pro Leu
305 310 315 320

Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly Val Leu Gly
325 330 335

Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu Phe Ser Thr
340 345 350

Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu Thr Leu Leu
355 360 365

Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly Lys Glu Val
370 375 380

Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Glu Asp Leu Pro Leu Pro
385 390 395 400

Glu Gly Ile Leu Leu Ala Leu Lys Gly Phe Ala Gly Lys Ile Arg Ser
405 410 415

Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu His Val Ser
420 425 430

Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro Thr Val Phe
435 440 445

Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val Trp Arg Glu
450 455 460

Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly Ala Ser Val
465 470 475 480

Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu Leu Trp Pro
485 490 495

Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val Ala Lys Met
500 505 510

Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala Ala Ala Phe
515 520 525

Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala Ala Val Leu
530 535 540

His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr Glu Gly Asn
545 550 555 560

Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro Leu Arg Glu
565 570 575

Ala Lys

<210> 5

<211> 832
<212> PRT
<213> myeloblastosis-associated virus

<220>
<223> beta (no met, no tag, no stop)

<400> 5

Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys Pro Asn His
1 5 10 15

Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly Lys Leu Val
20 25 30

Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly His Ile Glu
35 40 45

Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile Arg Lys Ala
50 55 60

Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val Asn Ala Lys
65 70 75 80

Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val Leu Ser Ala
85 90 95

Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys Asp Cys Phe
100 105 110

Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Arg Phe Ala Phe Thr
115 120 125

Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe Gln Trp Lys
130 135 140

Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys Gln Leu Ile
145 150 155 160

Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro Ser Leu Arg
165 170 175

Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser Ser His Asp
180 185 190

Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu Glu Arg Ala
195 200 205

Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Arg Glu Pro Gly Val Gln
210 215 220

Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Val Ala Pro Val Gly Leu
225 230 235 240

Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln Lys Leu Val
245 250 255

Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro Pro Arg Leu
260 265 270

Met Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro Asn Glu Ala
275 280 285

Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu Ile Val Gln
290 295 300

Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala Leu Pro Leu
305 310 315 320

Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly Val Leu Gly
325 330 335

Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu Phe Ser Thr
340 345 350

Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu Thr Leu Leu
355 360 365

Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly Lys Glu Val
370 375 380

Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Glu Asp Leu Pro Leu Pro
385 390 395 400

Glu Gly Ile Leu Leu Ala Leu Lys Gly Phe Ala Gly Lys Ile Arg Ser
405 410 415

Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu His Val Ser
420 425 430

Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro Thr Val Phe
435 440 445

Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val Trp Arg Glu
450 455 460

Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly Ala Ser Val
465 470 475 480

Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu Leu Trp Pro
485 490 495

Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val Ala Lys Met
500 505 510

Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala Ala Ala Phe
515 520 525

Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala Ala Val Leu
530 535 540

His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr Glu Gly Asn
545 550 555 560

Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro Leu Arg Glu
565 570 575

Ala Lys Asp Leu His Thr Ala Leu His Ile Gly Pro Arg Ala Leu Ser
580 585 590

Lys Ala Cys Asn Ile Ser Met Gln Gln Ala Arg Glu Val Val Gln Thr
595 600 605

Cys Pro His Cys Asn Ser Ala Pro Ala Leu Glu Ala Gly Val Asn Pro
610 615 620

Arg Gly Leu Gly Pro Leu Gln Ile Trp Gln Thr Asp Phe Thr Leu Glu
625 630 635 640

Pro Arg Met Ala Pro Arg Ser Trp Leu Ala Val Thr Val Asp Thr Ala
645 650 655

Ser Ser Ala Ile Val Val Thr Gln His Gly Arg Val Thr Ser Val Ala
660 665 670

Ala Gln His His Trp Ala Thr Ala Ile Ala Val Leu Gly Arg Pro Lys
675 680 685

Ala Ile Lys Thr Asp Asn Gly Ser Cys Phe Thr Ser Lys Ser Thr Arg
690 695 700

Glu Trp Leu Ala Arg Trp Gly Ile Ala His Thr Thr Gly Ile Pro Gly
705 710 715 720

Asn Ser Gln Gly Gln Ala Met Val Glu Arg Ala Asn Arg Leu Leu Lys
725 730 735

Asp Lys Ile Arg Val Leu Ala Glu Gly Asp Gly Phe Met Lys Arg Ile
740 745 750

Pro Thr Ser Lys Gln Gly Glu Leu Leu Ala Lys Ala Met Tyr Ala Leu
755 760 765

Asn His Phe Glu Arg Gly Glu Asn Thr Lys Thr Pro Ile Gln Lys His
770 775 780

Trp Arg Pro Thr Val Leu Thr Glu Gly Pro Pro Val Lys Ile Arg Ile
785 790 795 800

Glu Thr Gly Glu Trp Glu Lys Gly Trp Asn Val Leu Val Trp Gly Arg
805 810 815

Gly Tyr Ala Ala Val Lys Asn Arg Asp Thr Asp Lys Val Ile Trp Val
820 825 830

<210> 6
<211> 1734
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> alpha coding region (no met, no tag, no stop)

<400> 6
actgttgcgc tacatctggc tattccgctc aaatggaagc caaaccacac gcctgtgtgg 60

attgaccagt ggcccccttcc tgaaggtaaa cttgtagcgc taacgcaatt agtggaaaaa 120

gaattacagt taggacatat agaaccttca cttagttgct ggaacacacc tgtctttgtg 180

atccggaagg cttccgggtc ttatcgctta ttgcatgact tgcgcgctgt taacgctaag 240

cttgttcctt ttggggccgt ccaacagggg gcgccggttc tctccgcgct cccgcgtggt 300

tggcccctga tggtcctaga cctcaaggat tgcttctttt ctattcctct tgcggaacaa 360

gatcgcgaac gttttgcatt tacgctcccc tccgtgaata accaggcccc cgctcgaagg 420

ttccaatgga aggtcttgcc ccaagggatg acctgttctc ccactatctg tcagttgata 480

gtgggtcaaa tacttgagcc cttgcgactc aagcacccat ctctgcgcat gttgcattat 540

atggatgatc ttttgctagc cgcctcaagt catgatgggt tggaagcggc aggggaggag 600

gttatcagta cattggaaag agccgggttc accatttcgc ctgataaggt ccagagggag 660

cccggagtac aatatcttgg gtacaagtta ggtagtacgt atgtagcacc cgtaggcctg 720

gtagcagaac ccaggatagc caccttgtgg gatgttcaga agctggtggg gtcacttcag 780

tggcttcgcc cagcgttagg aatcccgcca cgactgatgg gccccttta tgagcagtta 840

cgagggtcag atcctaacga ggcgagggaa tggaatctag acatgaaaat ggcctggaga 900

gagatcgtac agctcagcac cactgctgcc ttggagcgat gggaccctgc cctgcctctg 960

gaaggagcgg tcgctagatg tgaacagggg gcaatagggg tcctgggaca gggactgtcc 1020

acacacccaa ggccatgttt gtggctattc tccacccaac ccaccaaggc gtttactgct 1080

tggttagaag tgctcaccct tttgattact aagctacgtg cttcggcagt gcgaaccttt 1140

ggcaaggagg ttgatatcct cctgttgcct gcatgctttc gggaggacct tccgctcccg 1200

gagggatcc tgttagccct taaggggttt gcaggaaaaa tcaggagtag tgacacgcca 1260

tctatttttg acattgcgcg tccactgcat gtttctctga aagtgagggt taccgaccac 1320

cctgtgccgg gacccactgt ctttactgac gcctcctcaa gcacccataa gggggtggta 1380

gtctaaagaa agaggcccaag gtgggagata aaagaaatag ctgatttggg ggcaagtgta 1440

caacaactgt aaagcaggc tatggccatg gcacttctgc tgtgggcgac aacgcccact 1500

aatgtaatga ctgactccgc gtttgttgcg aaaatgttac tcaagatggg acaggaggga 1560

gtcccgtcta cagcggcggc ttttatttta gaggatgcgt taagccaaag gtcagccatg 1620

gccgccgttc tccacgtgcg gagtcattct gaagtgccag ggtttttcac agaaggaaat 1680

gacgtggcag atagccaatc cacctttcaa gcgtatccct tgagagaggc taaa 1734

<210> 7
<211> 1737
<212> DNA
<213> myeloblastosis associated virus

<220>
<223> alpha coding region (no met, no tag, stop)

<400> 7
actgttgcgc tacatctggc tattccgctc aaatggaagc caaaccacac gcctgtgtgg 60

attgaccagt ggccccttcc tgaaggtaaa cttgtagcgc taacgcaatt agtggaaaaa 120

gaattacagt taggacatat agaaccttca cttagttgct ggaacacacc tgtctttgtg 180

atccggaagg cttccgggtc ttatcgctta ttgcatgact tgcgcgctgt taacgctaag 240

cttgttcctt ttggggccgt ccaacagggg gcgccggttc tctccgcgct cccgcgtggt 300

tggccctga tggtcctaga cctcaaggat tgcttctttt ctattcctct tgcggaacaa 360

gatcgcgaac gttttgcatt tacgctcccc tccgtgaata accaggcccc cgctcgaagg 420

ttccaatgga aggtcttgcc ccaagggatg acctgttctc ccactatctg tcagttgata 480

atgggtcaaa tacttgagcc cttgcgactc aagcacccat ctctgcgcat gttgcattat 540

atggatgatc ttttgctagc cgcctcaagt catgatgggt tggaagcggc aggggaggag 600

gttatcagta cattggaaag agccgggttc accatttcgc ctgataaggt ccagagggag 660

cccggagtac aatatcttgg gtacaagtta ggtagtacgt atgtagcacc cgtaggcctg 720

gtagcagaac ccaggatagc caccttgtgg gatgttcaga agctggtggg gtcacttcag 780

tggcttcgcc cagcgttagg aatcccgcca cgactgatgg gccccttttа tgagcagtta 840

cgagggtcag atcctaacga ggcgagggaa tggaatctag acatgaaaat ggcctggaga 900

gagatcgtac agctcagcac cactgctgcc ttggagcgat gggaccctgc cctgcctctg 960

gaaggagcgg tcgctagatg tgaacagggg gcaatagggg tcctgggaca gggactgtcc 1020

acacacccaa ggccatgttt gtggctattc tccacccaac ccaccaaggc gtttactgct 1080

tggttagaag tgctcaccct tttgattact aagctacgtg cttcggcagt gcgaaccttt 1140

ggcaaggagg ttgatatcct cctgttgcct gcatgctttc gggaggacct tccgctcccg 1200

gaggggatcc tgttagcccc taaggggttt gcaggaaaaa tcaggagtag tgacacgcca 1260

tctatttttg acattgcgcg tccactgcat gtttctctga aagtgagggt taccgaccac 1320

cctgtgccgg gacccactgt ctttactgac gcctcctcaa gcacccataa gggggtggta 1380

gtctggaggg agggcccaag gtgggagata aaagaaatag ctgatttggg ggcaagtgta 1440

caacaactgg aagcacgcgc tgtggccatg gcacttctgc tgtggccgac aacgcccact 1500

aatgtagtga ctgactccgc gtttgttgcg aaaatgttac tcaagatggg acaggaggga 1560

gtcccgtcta cagcggcggc ttttatttta gaggatgcgt taagccaaag gtcagccatg 1620

gccgccgttc tccacgtgcg gagtcattct gaagtgccag ggtttttcac agaaggaaat 1680

gacgtggcag atagccaagc cacctttcaa gcgtatccct tgagagaggc taaataa 1737

<210> 8
<211> 2496
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> beta coding region (no met, no tag, no stop)

<400> 8

actgttgcgc tacatctggc tattccgctc aaatggaagc caaaccacac gcctgtgtgg 60

attgaccagt ggccccttcc tgaaggtaaa cttgtagcgc taacgcaatt agtggaaaaa 120

gaattacagt taggacatat agaaccttca cttagttgct ggaacacacc tgtctttgtg 180

atccggaagg cttccgggtc ttatcgctta ttgcatgact tgcgcgctgt taacgctaag 240

cttgttcctt ttggggccgt ccaacagggg gcgccggttc tctccgcgct cccgcgtggt 300

tggcccctga tggtcctaga cctcaaggat tgcttctttt ctattcctct tgcggaacaa 360

gatcgcgaac gttttgcatt tacgctcccc tccgtgaata accaggcccc cgctcgaagg 420

ttccaatgga aggtcttgcc ccaagggatg acctgttctc ccactatctg tcagttgata 480

gtgggtcaaa tacttgagcc cttgcgactc aagcacccat ctctgcgcat gttgcattat 540

atggatgatc ttttgctagc cgcctcaagt catgatgggt tggaagcggc aggggaggag 600

gttatcagta cattggaaag agccgggttc accatttcgc ctgataaggt ccagagggag 660

cccggagtac aatatcttgg gtacaagtta ggtagtacgt atgtagcacc cgtaggcctg 720

gtagcagaac ccaggatagc caccttgtgg gatgttcaga agctggtggg gtcacttcag 780

tggcttcgcc cagcgttagg aatcccgcca cgactgatgg gccccttttа tgagcagtta 840

cgagggtcag atcctaacga ggcgagggaa tggaatctag acatgaaaat ggcctggaga 900

gagatcgtac agctcagcac cactgctgcc ttggagcgat gggaccctgc cctgcctctg 960

gaaggagcgg tcgctagatg tgaacagggg gcaatagggg tcctgggaca gggactgtcc 1020

acacacccaa ggccatgttt gtggctattc tccacccaac ccaccaaggc gtttactgct 1080

tggttagaag tgctcaccct tttgattact aagctacgtg cttcggcagt gcgaaccttt 1140

ggcaaggagg ttgatatcct cctgttgcct gcatgctttc gggaggacct tccgctcccg 1200

gaggggatcc tgttagcccc taaggggttt gcaggaaaaa tcaggagtag tgacacgcca 1260

tctatttttg acattgcgcg tccactgcat gtttctctga aagtgagggt taccgaccac 1320

cctgtgccgg gatccactgt ctttactgac gcctcctcaa gcacccataa gggggtggta 1380

gtctgggggt gtttccaaag gtgggagata aaagaaatag ctgatttggg ggcaagtgta 1440

caacaactgg aagcatgcgc tgtggccatg gcacttctgc tgtggccgac aacgcccact 1500

aatgtagtga ctgactccgc gtttgttgcg aaaatgttac tcaagatggg acaggaggga 1560

gtcccgtcta cagcggcggc ttttatttta gaggatgcgt taagccaaag gtcagccatg 1620

gccgccgttc tccatgtgcg gagtcattct gaagtgccag ggtttttcac agaaggaaat 1680

gacgtggcag atagccaagc cacctttcaa gcgtatccct tgagagaggc taaagatctc 1740

cataccgctc tccatattgg accccgcgcg ctatccaaag cgtgtaatat atctatgcag 1800

caggctaggg aggttgttca gacctgcccg cattgtaatt cagcccctgc gttggaggcc 1860

ggggtaaacc ctagggqgttt gggacccccta cagatatggc agacagactt tacacttgag 1920

cctagaatgg cccccccgttc ctggctcgct gttactgtgg ataccgcctc atcggcgata 1980

gtcgtaactc agcatggccg tgtcacatcg gttgctgcac aacatcattg ggccacggct 2040

atcgccgttt tgggaagacc aaaggccata aaaacagata atgggtcctg cttcacgtct 2100

aaatccacgc gagagtggct cgcgagatgg gggatagcac acaccaccgg gattccgggt 2160

aattcccagg gtcaagctat ggtagagcgg gccaaccggc tcctgaaaga taagatccgt 2220

gtgcttgcgg agggggatgg ctttatgaaa agaatcccca ccagcaaaca gggggaacta 2280

ttagccaagg caatgtatgc cctcaatcac tttgagcgtg gtgaaaacac aaaaacaccg 2340

atacaaaaac actggagacc taccgttctt acagaaggac ccccggttaa aatacgaata 2400

gagacagggg agtgggaaaa aggatggaac gtgctggtct ggggacgagg ttatgccgct 2460

gtgaaaaaca gggacactga taaggttatt tgggta 2496

<210> 9
<211> 2499
<212> DNA

<220> myeloblastosis associated virus

<220>

<221> beta coding region (no met, no tag, stop)

<400> 9

atttgcgc tacatctggc tattccgctc aaatggaagc caaaccacac gcctgtgtgg 60

attaaccagt ggccccttcc tgaaggtaaa cttgtagcgc taacgcaatt agtggaaaaa 120

agattacagt taggacatat agaaccttca cttagttgct ggaacacacc tgtctttgtg 180

attggaaagg cttccgggtc ttatcgctta ttgcatgact tgcgcgctgt taacgctaag 240

cttgttcctt ttgggggcgt ccaacagggg gcgccggttc tctccgcgct cccgcgtggt 300

tggcccctga tggtcctaga cctcaaggat tgcttctttt ctattcctct tgcggaacaa 360

gattggcgaac gttttgcatt tacgctcccc tccgtgaata accaggcccc cgctcgaagg 420

ttccaatgga aggtcttgcc ccaagggatg acctgttctc ccactatctg tcagttgata 480

gtggttcaaa tacttgagcc cttgcgactc aagcacccat ctctgcgcat gttgcattat 540

atgaatgatc ttttgctagc cgcctcaagt catgatgggt tggaagcggc aggggaggag 600

gttatcagta cattggaaag agccgggttc accatttcgc ctgataaggt ccagagggag 660

ccctggagtac aatatcttgg gtacaagtta ggtagtacgt atgtagcacc cgtaggcctg 720

gtatgcagaac ccaggatagc caccttgtgg gatgttcaga agctggtggg gtcacttcag 780

tgttttcgcc cagcgttagg aatcccgcca cgactgatgg gccccttta tgagcagtta 840

cgaaggtcag atcctaacga ggcgagggaa tggaatctag acatgaaaat ggcctggaga 900

gaaatcgtac agctcagcac cactgctgcc ttggagcgat gggaccctgc cctgcctctg 960

gaacgagcgg tcgctagatg tgaacagggg gcaatagggg tcctgggaca gggactgtcc 1020

acacacccaa ggccatgttt gtggctattc tccacccaac ccaccaaggc gtttactgct 1080

tggttagaag tgctcaccct tttgattact aagctacgtg cttcggcagt gcgaaccttt 1140

ggcaaggagg ttgatatcct cctgttgcct gcatgctttc gggaggacct tccgctcccg 1200

gaggggatcc tgttagcact taaggggttt gcaggaaaaa tcaggagtag tgacacgcca 1260

tctatttttg acattgcgcg tccactgcat gtttctctga aagtgagggt taccgaccac 1320

cctgtgccgg gacccactgt ctttactgac gcctcctcaa gcacccataa gggggtggta 1380

gtctggaggg agggcccaag gtgggagata aaagaaatag ctgatttggg ggcaagtgta 1440

caacaactgg aagcacgcgc tgtggccatg gcacttctgc tgtggccgac aacgcccact 1500

aatgtagtga ctgactccgc gtttgttgcg aaaatgttac tcaagatggg acaggaggga 1560

gtcccgtcta cagcggcggc ttttatttta gaggatgcgt taagccaaag gtcagccatg 1620

gccgccgttc tccacgtgcg gagtcattct gaagtgccag ggtttttcac agaaggaaat 1680

gacgtggcag atagccaagc cacctttcaa gcgtatccct tgagagaggc taaagatctc 1740

cataccgctc tccatattgg accccgcgcg ctatccaaag cgtgtaatat atctatgcag 1800

caggctaggg aggttgttca gacctgcccg cattgtaatt cagcccctgc gttggaggcc 1860

ggggtaaacc ctaggggttt gggaccccta cagatatggc agacagactt tacacttgag 1920

cctagaatgg ccccccgttc ctggctcgct gttactgtgg ataccgcctc atcggcgata 1980

gtcgtaactc agcatggccg tgtcacatcg gttgctgcac aacatcattg ggccacggct 2040

atcgccgttt tgggaagacc aaaggccata aaaacagata atgggtcctg cttcacgtct 2100

aaatccacgc gagagtggct cgcgagatgg gggatagcac acaccaccgg gattccgggt 2160

aattcccagg gtcaagctat ggtagagcgg gccaaccggc tcctgaaaga taagatccgt 2220

gtgcttgcgg agggggatgg ctttatgaaa agaatcccca ccagcaaaca gggggaacta 2280

ttagccaagg caatgtatgc cctcaatcac tttgagcgtg gtgaaaacac aaaaacaccg 2340

atacaaaaac actggagacc taccgttctt acagaaggac ccccggttaa aatacgaata 2400

gagacagggg agtgggaaaa aggatggaac gtgctggtct ggggacgagg ttatgccgct 2460

gtgaaaaaca gggacactga taaggttatt tgggtataa 2499

<210> 10
<211> 2688
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> full-length coding region (no met, no tag, stop)

<400> 10
actgttgcgc tacatctggc tattccgctc aaatggaagc caaaccacac gcctgtgtgg 60

attgaccagt ggccccttcc tgaaggtaaa cttgtagcgc taacgcaatt agtggaaaaa 120

gaattacagt taggacatat agaaccttca cttagttgct ggaacacacc tgtctttgtg 180

atccggaagg cttccgggtc ttatcgctta ttgcatgact tgcgcgctgt taacgctaag 240

cttgttcctt ttggggccgt ccaacagggg gcgccggttc tctccgcgct cccgcgtggt 300

tggcccctga tggtcctaga cctcaaggat tgcttctttt ctattcctct tgcggaacaa 360

gatcgcgaac gttttgcatt tacgctcccc tccgtgaata accaggcccc cgctcgaagg 420

ttccaatgga aggtcttgcc ccaagggatg acctgttctc ccactatctg tcagttgata 480

gtgggtcaaa tacttgagcc cttgcgactc aagcacccat ctctgcgcat gttgcattat 540

atggatgatc ttttgctagc cgcctcaagt catgatgggt tggaagcggc aggggaggag 600

gttatcagta cattggaaag agccgggttc accatttcgc ctgataaggt ccagagggag 660

cccggagtac aatatcttgg gtacaagtta ggtagtacgt atgtagcacc cgtaggcctg 720

gtagcagaac ccaggatagc caccttgtgg gatgttcaga agctggtggg gtcacttcag 780

tggcttcgcc cagcgttagg aatcccgcca cgactgatgg gccccttta tgagcagtta 840

cgagggtcag atcctaacga ggcgagggaa tggaatctag acatgaaaat ggcctggaga 900

gagatcgtac agctcagcac cactgctgcc ttggagcgat gggaccctgc cctgcctctg 960

gaagtaqcaa tcgctagatg tgaacagggg gcaatagggg tcctgggaca gggactgtcc 1020

agatatggaa ggccatgttt gtggctattc tccacccaac ccaccaaggc gtttactgct 1080

tggttagaag tgctcaccct tttgattact aagctacgtg cttcggcagt gcgaaccttt 1140

ggtaaggaag ttgatatcct cctgttgcct gcatgctttc gggaggacct tccgctcccg 1200

gaggtgatcc tgttagccct taaggggttt gcaggaaaaa tcaggagtag tgacacgcca 1260

tctatttttg acattgcgcg tccactgcat gtttctctga aagtgagggt taccgaccac 1320

cctatgccgg gacccactgt ctttactgac gcctcctcaa gcacccataa gggggtggta 1380

gtctaggaggg agggcccaag gtgggagata aaagaaatag ctgatttggg ggcaagtgta 1440

caacaactgg aagcacgcgc tgtggccatg gcacttctgc tgtggccgac aacgcccact 1500

aatgtagtga ctgactccgc gtttgttgcg aaaatgttac tcaagatggg acaggaggga 1560

gtcccgtcta cagcggcggc ttttatttta gaggatgcgt taagccaaag gtcagccatg 1620

gccgccgttc tccacgtgcg gagtcattct gaagtgccag ggtttttcac agaaggaaat 1680

gacgtggcag atagccaagc cacctttcaa gcgtatccct tgagagaggc taaagatctc 1740

catatccgct tccatattgg accccgcgcg ctatccaaag cgtgtaatat atctatgcag 1800

caggctaggg aggttgttca gacctgcccg cattgtaatt cagcccctgc gttggaggcc 1860

gtggtaaacc ctaggggttt gggacccctа cagatatggc agacagactt tacacttgag 1920

ctagaatgg ccccccgttc ctggctcgct gttactgtgg ataccgcctc atcggcgata 1980

gtcgtaactc agcatggccg tgtcacatcg gttgctgcac aacatcattg ggccacggct 2040

atcgccgttt tgggaagacc aaaggccata aaaacagata atgggtcctg cttcacgtct 2100

aaatccacgc gagagtggct cgcgagatgg gggatagcac acaccaccgg gattccgggt 2160

aattcccagg gtcaagctat ggtagagcgg gccaaccggc tcctgaaaga taagatccgt 2220

gtgcttgcgg aggggatgg ctttatgaaa agaatcccca ccagcaaaca gggggaacta 2280

ttagccaagg caatgtatgc cctcaatcac tttgagcgtg gtgaaaacac aaaaacaccg 2340

atacaaaaac agtggagacc taccgttctt acagaaggac ccccggttaa aatacgaata 2400

gagacagggg agtgggaaaa aggatggaac gtgctggtct ggggacgagg ttatgccgct 2460

gtgaaaaaca gggacactga taaggttatt tgggtaccct ctcgaaaagt taaaccggac 2520

atcacccaaa aggatgaggt gactaagaaa gatgaggcga gccctctttt tgcaggcatt 2580

tctgactggg cgccctggga aggcgagcaa gaaggactcc aagaagaaac cgccagcaac 2640

aagcaagaaa gacccggaga agacaccccc gctgccaacg agagttaa 2688

<210> 11
<211> 2691
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> full-length coding region (met, no tag, stop)

<400> 11
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccc tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gacccggaag tacaatatct tgggtacaaa ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gggaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa accctagggg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggccccccg ttcctggctc gctgttactg tggataccgc ctcatcgggg 1980

atagtcgtaa ctcagcatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatcgccg ttttgggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaatcca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

ggtaattccc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

cgtgtgcttg cggagggggа tggctttatg aaaagaatcc ccaccagcaa acaggggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

ccgatacaaa aacactggag acctaccgtt cttacagaag gacccccggt taaaatacga 2400

atagagacag gggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggtac cctctcgaaa agttaaaccg 2520

gacatcaccc aaaaggatga ggtgactaag aaagatgagg cgagccctct ttttgcaggc 2580

atttctgact gggcgccctg ggaaggcgag caagaaggac tccaagaaga aaccgccagc 2640

aacaagcaag aaagacccgg agaagacacc cctgctgcca acgagagtta a 2691

<210> 12
<211> 2499
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> beta coding region (met, no tag, no stop)

<400> 12
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccc tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttgggge cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctatttcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taagggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

gaagtccgtt ctatagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atgaccgaag ttttgaaaat gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagataagca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gaaaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa agttaaaagg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggaaaaatg ttcctggctc gctgttactg tggataccgc ctcatcggcg 1980

atagtcgtaa ctcaacatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatcgccg ttttgggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaatcca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

ggtaattccc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

cgtgtgcttg cggagggggg tggctttatg aaaagaatcc ccaccagcaa acaggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

ccgatacaaa aacactggag acctaccgtt cttacagaag gacccccggt taaaatacga 2400

atagagacag gggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggta 2499

<210> 13
<211> 1737
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> alpha coding region (met, no tag, no stop)

<400> 13
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggcccct tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taagggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc agggtttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaa 1737

<210> 14
<211> 2706
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> full-length coding region (met, his tag, no stop)

<400> 14
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccc tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggga tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcaa aaccaagagat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcactag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gggaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa accctagggg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggccccccg ttcctggctc gctgttactg tggataccgc ctcatcggcg 1980

atagtcgtaa ctcagcatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatgaccg ttttgcggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaatcca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

gttaattttc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

catgtgcttg cggaggggga tggctttatg aaaagaatcc ccaccagcaa acaggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

gctatacaaa aacactggag acctaccgtt cttacagaag gacccccggt taaaatacga 2400

atagaggacag aggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggtac cctctcgaaa agttaaaccg 2520

gatatcaccc aaaaggatga ggtgactaag aaagatgagg cgagccctct ttttgcaggc 2580

atttctgact gggcgccctg ggaaggcgag caagaaggac tccaagaaga aaccgccagc 2640

aacaagcaag aaagacccgg agaagacacc cctgctgcca acgagagtca ccaccaccac 2700

caccac 2706

<210> 15
<211> 2517
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> beta coding region (met, his tag, no stop)

<400> 15
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccct tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc ctttgggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tctttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atgaccgccg ttctccacgt gcggagtcat tctgaaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gggaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa accctagggg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggcccccccg ttcctggctc gctgttactg tggataccgc ctcatcggcg 1980

atagtcgtaa ctcagcatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatcgccg ttttgggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaatcca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

ggtaattccc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

cgtgtgcttg cggaggggga tggctttatg aaaagaatcc ccaccagcaa acaggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

ccgatacaaa aacactggag acctaccgtt cttacagaag gacccccggt taaaatacga 2400

atagaagacag gggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggtac accaccacca ccaccac 2517

<210> 16
<211> 1755
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> alpha coding region (met, his tag, no stop)

<400> 16
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tagattgacc agtggccgct tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caaaatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

agcttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgaaggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

cacccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtgggc atggcacttc tgctgtgggc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaacac 1740

caccaccacc accac 1755

<210> 17
<211> 2709
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> full-length coding region (met, his tag, stop)

<400> 17
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccc tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gggaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa accctagggg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggccccccg ttcctggctc gctgttactg tggataccgc ctcatcggcg 1980

atagtcgtaa ctcagcatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatcgccg ttttgggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaattca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

ggtaattccc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

cgtgtgcttg cggaggggga tggctttatg aaaagaatcc ccaccagcaa acaggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

ccgatacaaa aacactggag acctaccgtt cttacagaag gacccccggt taaaatacga 2400

atagagacag gggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggtac cctctcgaaa agttaaaccg 2520

gacatcaccc aaaaggatga ggtgactaag aaagatgagg cgagccctct ttttgcaggc 2580

atttctgact gggcgccctg ggaaggcgag caagaaggac tccaagaaga aaccgccagc 2640

aacaagcaag aaagacccgg agaagacacc cctgctgcca acgagagtca ccaccaccac 2700

caccactaa 2709

<210> 18
<211> 2520
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> beta coding region (met, his tag, stop)

<400> 18
atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccccc tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc ctttgggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

caccctgtgc cgggacccac tgtctttact gacgcctcct caagcaccca taaggggtg 1380

gtagtctgga gggagggccc aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tggaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggaagtcccgt ctacagcggc ggcttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

atggccgccg ttctccacgt gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg cagatagcca agccaccttt caagcgtatc ccttgagaga ggctaaagat 1740

ctccataccg ctctccatat tggaccccgc gcgctatcca aagcgtgtaa tatatctatg 1800

cagcaggcta gggaggttgt tcagacctgc ccgcattgta attcagcccc tgcgttggag 1860

gccggggtaa accctagggg tttgggaccc ctacagatat ggcagacaga ctttacactt 1920

gagcctagaa tggcccccccg ttcctggctc gctgttactg tggataccgc ctcatcggcg 1980

atagtcgtaa ctcagcatgg ccgtgtcaca tcggttgctg cacaacatca ttgggccacg 2040

gctatcgccg ttttgggaag accaaaggcc ataaaaacag ataatgggtc ctgcttcacg 2100

tctaaatcca cgcgagagtg gctcgcgaga tgggggatag cacacaccac cgggattccg 2160

ggtaattccc agggtcaagc tatggtagag cgggccaacc ggctcctgaa agataagatc 2220

cgtgtgcttg cggaggggga tggctttatg aaaagaatcc ccaccagcaa acaggggaa 2280

ctattagcca aggcaatgta tgccctcaat cactttgagc gtggtgaaaa cacaaaaaca 2340

ccgatacaaa aacactggag acctaccgtt cttacaguag gacccccggt taaaatacga 2400

atagagacag gggagtggga aaaaggatgg aacgtgctgg tctggggacg aggttatgcc 2460

gctgtgaaaa acagggacac tgataaggtt atttgggtac accaccacca ccaccactaa 2520

<210> 19
<211> 1758
<212> DNA
<213> myeloblastosis-associated virus

<220>
<223> alpha coding region (met, his tag, stop)

<400> 19

atgactgttg cgctacatct ggctattccg ctcaaatgga agccaaacca cacgcctgtg 60

tggattgacc agtggccgct tcctgaaggt aaacttgtag cgctaacgca attagtggaa 120

aaagaattac agttaggaca tatagaacct tcacttagtt gctggaacac acctgtcttt 180

gtgatccgga aggcttccgg gtcttatcgc ttattgcatg acttgcgcgc tgttaacgct 240

aagcttgttc cttttggggc cgtccaacag ggggcgccgg ttctctccgc gctcccgcgt 300

ggttggcccc tgatggtcct agacctcaag gattgcttct tttctattcc tcttgcggaa 360

caagatcgcg aacgttttgc atttacgctc ccctccgtga ataaccaggc ccccgctcga 420

aggttccaat ggaaggtctt gccccaaggg atgacctgtt ctcccactat ctgtcagttg 480

atagtgggtc aaatacttga gcccttgcga ctcaagcacc catctctgcg catgttgcat 540

tatatggatg atcttttgct agccgcctca agtcatgatg ggttggaagc ggcaggggag 600

gaggttatca gtacattgga aagagccggg ttcaccattt cgcctgataa ggtccagagg 660

gagcccggag tacaatatct tgggtacaag ttaggtagta cgtatgtagc acccgtaggc 720

ctggtagcag aacccaggat agccaccttg tgggatgttc agaagctggt ggggtcactt 780

cagtggcttc gcccagcgtt aggaatcccg ccacgactga tgggcccctt ttatgagcag 840

ttacgagggt cagatcctaa cgaggcgagg gaatggaatc tagacatgaa aatggcctgg 900

agagagatcg tacagctcag caccactgct gccttggagc gatgggaccc tgccctgcct 960

ctggaaggag cggtcgctag atgtgaacag ggggcaatag gggtcctggg acagggactg 1020

tccacacacc caaggccatg tttgtggcta ttctccaccc aacccaccaa ggcgtttact 1080

gcttggttag aagtgctcac ccttttgatt actaagctac gtgcttcggc agtgcgaacc 1140

tttggcaagg aggttgatat cctcctgttg cctgcatgct ttcgggagga ccttccgctc 1200

ccggagggga tcctgttagc ccttaagggg tttgcaggaa aaatcaggag tagtgacacg 1260

ccatctattt ttgacattgc gcgtccactg catgtttctc tgaaagtgag ggttaccgac 1320

[illegible] 1380

gtaatctgga [illegible] aaggtgggag ataaaagaaa tagctgattt gggggcaagt 1440

gtacaacaac tgaaagcacg cgctgtggcc atggcacttc tgctgtggcc gacaacgccc 1500

actaatgtag tgactgactc cgcgtttgtt gcgaaaatgt tactcaagat gggacaggag 1560

ggagtccegt [illegible] ggctttttatt ttagaggatg cgttaagcca aaggtcagcc 1620

at[illegible] ttct [illegible] gcggagtcat tctgaagtgc cagggttttt cacagaagga 1680

aatgacgtgg [illegible] aggccacctt caagcgtatc ccttgagaga ggctaaacac 1740

caccaccacc accactaa 1758

<210> 20
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: FSDRT

<400> 20
tgtactaagg aggtgttcat gactgttgcg ctacatctgg ct 42

<210> 21
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: RSDBAC2

<400> 21
gccagatgta gcgcaacagt catatttata ggttttttta ttac 44

<210> 22
<211> 20
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: FRT

<400> 22
atgactgttg cgctacatct 20

<210> 23
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: M1BARSDHIS

<400> 23
acccggatca attaattagt ggtggtggtg gtggtgttta gcctctctca agggata 57

<210> 24
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: M1KARSDHIS

<400> 24
acccggatca attaattagt ggtggtggtg gtggtgccaa ataaccttat cagt 54

<210> 25
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: FM1BASmaI

<400> 25
ataagggcca ctgttctccc cgggatgact gttgcgctgc atct 44

<210> 26
<400> 26

<210> 27
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer-M1BA

<400> 27
tttagcctct ctcaagggat a 21

<210> 28
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer-M1KA

<400> 28
taccccaaata accttatcag t 21

<210> 29
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: forward sequencing primer or FSP

<400> 29
cgccagggtt ttcccagtca cga 23

<210> 30
<211> 32
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: gene-specific capture primer

<400> 30
aaactatgcc aactagagat tggaggttgt tt 32

<210> 31
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amplification primer 1

<400> 31
accccatcca atgcatgtct cgggtcgtag tcttaaccat 40

<210> 32
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amplification primer 2

<400> 32
cgattccgct ccagacttct cgggtgctga aggagtaagg 40

<210> 33
<211> 23
<212> DNA
<213> Unknown

<400> 33
ggccacagca ccaccaccac cac 23

<210> 34
<211> 23
<212> DNA
<213> RNhis

<400> 34
ggccgtggtg gtggtggtgg tgt 23

<210> 35
<211> 64
<212> DNA
<213> RM1KAhisKpnI

<400> 35
tttaactttt cgagagggta ccttagtggt ggtggtggtg gtgtacccaa ataaccttat 60
cagt 64

<210> 36
<211> 66
<212> DNA
<213> RM1AAhisAccI

<400> 36
aaaataaaag ccgccgctgt cgacttagtg gtggtggtgg tggtggactc ctcctgtcc 60
catctt 66

<210> 37
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HRP-conjugated
P2 comp

<400> 37
ccttactcct tcagcacccg agaagtctgg agcggaatcg 40

<210> 38
<211> 3168
<212> DNA
<213> Human immunodeficiency virus type 2

<220>
<221> CDS
<222> (1)..(3168)

<400> 38

atg ctg gaa atg tgg aca gca agg aca cat cat gtc aaa atg ccc aga 48
Met Leu Glu Met Trp Thr Ala Arg Thr His His Val Lys Met Pro Arg
1 5 10 15

aag aca ggc ggg ttt ttt agg gtt cgg ccc ctg ggg aaa gaa gcc tcg 96
Lys Thr Gly Gly Phe Phe Arg Val Arg Pro Leu Gly Lys Glu Ala Ser
20 25 30

caa ttt ccc cgt cca ggc acc cca ggg gat agt gcc atc tgc gcc ccc 144
Gln Phe Pro Arg Pro Gly Thr Pro Gly Asp Ser Ala Ile Cys Ala Pro
35 40 45

gat gaa ccc agc att cgg cat gac acc tca ggg tgc gat tcc atc tgc 192
Asp Glu Pro Ser Ile Arg His Asp Thr Ser Gly Cys Asp Ser Ile Cys
50 55 60

acc ccc tgc aga tcc agc aga gga gat gct aaa gaa cta cat gca act 240
Thr Pro Cys Arg Ser Ser Arg Gly Asp Ala Lys Glu Leu His Ala Thr
65 70 75 80

agg gaa gaa gca gaa gga gaa cag aga gag acc cta caa gga ggt gac 288
Arg Glu Glu Ala Glu Gly Glu Gln Arg Glu Thr Leu Gln Gly Gly Asp
85 90 95

aga gga ttt gct gca cct caa ttc tct ctt tgg aga aga cca gta gtc 336
Arg Gly Phe Ala Ala Pro Gln Phe Ser Leu Trp Arg Arg Pro Val Val
100 105 110

aaa gca act att gag ggt caa tca gta gaa gta tta cta gac aca gga 384
Lys Ala Thr Ile Glu Gly Gln Ser Val Glu Val Leu Leu Asp Thr Gly
115 120 125

gct gat gac tca ata gta gca ggg ata gaa tta ggc agc aat tac acc 432
Ala Asp Asp Ser Ile Val Ala Gly Ile Glu Leu Gly Ser Asn Tyr Thr

130 135 140

cca aaa ata gta ggt ggg ata gga gga ttt ata aat acc aat gaa tac 480
Pro Lys Ile Val Gly Gly Ile Gly Gly Phe Ile Asn Thr Asn Glu Tyr
145 150 155 160

aaa aat gta gaa ata gaa gta gta gga aaa aga gta aga gca aca gta 528
Lys Asn Val Glu Ile Glu Val Val Gly Lys Arg Val Arg Ala Thr Val
165 170 175

atg aca ggg gac acc cca ata aac att ttt ggc aga aat att tta aat 576
Met Thr Gly Asp Thr Pro Ile Asn Ile Phe Gly Arg Asn Ile Leu Asn
180 185 190

agc tta ggc atg act cta aat ttc cca gta gca agg ata gaa cca gta 624
Ser Leu Gly Met Thr Leu Asn Phe Pro Val Ala Arg Ile Glu Pro Val
195 200 205

aaa gtc cag tta aag cct gaa aaa gat ggg cca aaa atc aga caa tgg 672
Lys Val Gln Leu Lys Pro Glu Lys Asp Gly Pro Lys Ile Arg Gln Trp
210 215 220

ccc cta tcc aaa gag aaa ata cta gcc ctc aaa gaa atc tgt gaa aaa 720
Pro Leu Ser Lys Glu Lys Ile Leu Ala Leu Lys Glu Ile Cys Glu Lys
225 230 235 240

atg gaa aaa gag gga cag tta gaa gag gcg cct cct act aat cca tac 768
Met Glu Lys Glu Gly Gln Leu Glu Glu Ala Pro Pro Thr Asn Pro Tyr
245 250 255

aat tcg ccc acc ttc gcc ata aaa aag aaa gac aaa aac aaa tgg agg 816
Asn Ser Pro Thr Phe Ala Ile Lys Lys Lys Asp Lys Asn Lys Trp Arg
260 265 270

atg cta ata gat ttc aga gaa cta aac aag gta acc caa gaa ttt aca 864
Met Leu Ile Asp Phe Arg Glu Leu Asn Lys Val Thr Gln Glu Phe Thr
275 280 285

gag gtc cag ctg ggt att cct cac cca gca gga ctg gca tca aag aaa 912
Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Ala Ser Lys Lys
290 295 300

aga ata aca gta cta gat gta gga gat gcc tac ttc agt gtc cca cta 960
Arg Ile Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu

305 310 315 320

gat cca gac ttc aga caa tat aca gca ttt act ttg cca gca gta aat 1008
Asp Pro Asp Phe Arg Gln Tyr Thr Ala Phe Thr Leu Pro Ala Val Asn
325 330 335

aat gca gaa cca gga aag aga tat ctt tac aaa gtc cta cca cag gga 1056
Asn Ala Glu Pro Gly Lys Arg Tyr Leu Tyr Lys Val Leu Pro Gln Gly
340 345 350

tgg aag gga tcc cca gca att ttc cag tac acc atg gca aag gta cta 1104
Trp Lys Gly Ser Pro Ala Ile Phe Gln Tyr Thr Met Ala Lys Val Leu
355 360 365

gac cct ttc aga aaa gcc aac aat gat gtc act ata atc cag tac atg 1152
Asp Pro Phe Arg Lys Ala Asn Asn Asp Val Thr Ile Ile Gln Tyr Met
370 375 380

gat gac att ctc gtg gca agt gac agg agc gat ctg gag cat gac agg 1200
Asp Asp Ile Leu Val Ala Ser Asp Arg Ser Asp Leu Glu His Asp Arg
385 390 395 400

gta gtg tct caa cta aaa gag cta tta aat aac atg gga ttc tct act 1248
Val Val Ser Gln Leu Lys Glu Leu Leu Asn Asn Met Gly Phe Ser Thr
405 410 415

cca gaa gaa aag ttc caa aaa gac cct cca ttc aaa tgg atg ggg tat 1296
Pro Glu Glu Lys Phe Gln Lys Asp Pro Pro Phe Lys Trp Met Gly Tyr
420 425 430

gag ctc tgg cca aag aaa tgg aaa ctg caa aaa ata cag cta cca gaa 1344
Glu Leu Trp Pro Lys Lys Trp Lys Leu Gln Lys Ile Gln Leu Pro Glu
435 440 445

aaa gag gtt tgg aca gta aat gac att cag aag tta gtg gga gta tta 1392
Lys Glu Val Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Val Leu
450 455 460

aat tgg gca gct caa ctt ttc ccg ggg att aag acc agg cat ata tgt 1440
Asn Trp Ala Ala Gln Leu Phe Pro Gly Ile Lys Thr Arg His Ile Cys
465 470 475 480

aaa cta ata agg gga aag atg acc cta aca gaa gag gta caa tgg act 1488
Lys Leu Ile Arg Gly Lys Met Thr Leu Thr Glu Glu Val Gln Trp Thr

485 490 495

gaa ttg gca gag gca gaa ttc cag gaa aac aaa atc atc cta gaa caa 1536
Glu Leu Ala Glu Ala Glu Phe Gln Glu Asn Lys Ile Ile Leu Glu Gln
500 505 510

gag cag gaa gga tcc tat tac aaa gaa ggg gta cct tta gaa gca aca 1584
Glu Gln Glu Gly Ser Tyr Tyr Lys Glu Gly Val Pro Leu Glu Ala Thr
515 520 525

gtg cag aaa aat cta gca aat cag tgg aca tac aag att cat cag gga 1632
Val Gln Lys Asn Leu Ala Asn Gln Trp Thr Tyr Lys Ile His Gln Gly
530 535 540

gat aaa atc cta aaa gta gga aaa tat gca aag gtt aaa aac act cac 1680
Asp Lys Ile Leu Lys Val Gly Lys Tyr Ala Lys Val Lys Asn Thr His
545 550 555 560

acc aat gga gta aga cta ttg gct cat gta gtc caa aaa ata gga aag 1728
Thr Asn Gly Val Arg Leu Leu Ala His Val Val Gln Lys Ile Gly Lys
565 570 575

gaa gca ttg gtc atc tgg gga gag ata cca atg ttc cat cta cca gta 1776
Glu Ala Leu Val Ile Trp Gly Glu Ile Pro Met Phe His Leu Pro Val
580 585 590

gaa aga gag aca tgg gat cag tgg tgg aca gat tac tgg caa gta acc 1824
Glu Arg Glu Thr Trp Asp Gln Trp Trp Thr Asp Tyr Trp Gln Val Thr
595 600 605

tgg atc cca gaa tgg gat ttt gtc tca acc cca cca tta ata agg tta 1872
Trp Ile Pro Glu Trp Asp Phe Val Ser Thr Pro Pro Leu Ile Arg Leu
610 615 620

gcc tat aac ctg gtc aaa gac ccc cta gaa gga gta gaa act tac tac 1920
Ala Tyr Asn Leu Val Lys Asp Pro Leu Glu Gly Val Glu Thr Tyr Tyr
625 630 635 640

aca gat gga tcc tgt aac aaa gcc tca aaa gaa ggg aaa gca gga tat 1968
Thr Asp Gly Ser Cys Asn Lys Ala Ser Lys Glu Gly Lys Ala Gly Tyr
645 650 655

gtc aca gac agg gga aag gat aaa gtt aaa cca tta gaa caa aca aca 2016
Val Thr Asp Arg Gly Lys Asp Lys Val Lys Pro Leu Glu Gln Thr Thr

660 665 670

aat cag caa gca gag ctt gaa gca ttt gca cta gca cta cag gac tca 2064
Asn Gln Gln Ala Glu Leu Glu Ala Phe Ala Leu Ala Leu Gln Asp Ser
675 680 685

gga cca cag gtc aat atc ata gta gat tca caa tat gtc atg gga ata 2112
Gly Pro Gln Val Asn Ile Ile Val Asp Ser Gln Tyr Val Met Gly Ile
690 695 700

gta gct gca caa cca aca gaa aca gaa tca ccg ata gta aga gaa ata 2160
Val Ala Ala Gln Pro Thr Glu Thr Glu Ser Pro Ile Val Arg Glu Ile
705 710 715 720

att gaa gaa atg atc aaa aag gaa aaa ata tat gta gga tgg gta cca 2208
Ile Glu Glu Met Ile Lys Lys Glu Lys Ile Tyr Val Gly Trp Val Pro
725 730 735

gct cac aag gga ctg ggt ggt aat cag gaa gta gac cac cta gtg agc 2256
Ala His Lys Gly Leu Gly Gly Asn Gln Glu Val Asp His Leu Val Ser
740 745 750

caa gga att aga caa atc cta ttt cta gaa aaa ata gaa cca gct caa 2304
Gln Gly Ile Arg Gln Ile Leu Phe Leu Glu Lys Ile Glu Pro Ala Gln
755 760 765

gaa gaa cat gaa aaa tat cat aat aat gta aaa gaa cta gtc cat aaa 2352
Glu Glu His Glu Lys Tyr His Asn Asn Val Lys Glu Leu Val His Lys
770 775 780

ttt ggg att cca caa tta gtg gca aga caa ata gta aat tcc tgt gat 2400
Phe Gly Ile Pro Gln Leu Val Ala Arg Gln Ile Val Asn Ser Cys Asp
785 790 795 800

aaa tgc caa caa aaa ggg gaa gct att cat gga cag gta aat tca gaa 2448
Lys Cys Gln Gln Lys Gly Glu Ala Ile His Gly Gln Val Asn Ser Glu
805 810 815

cta ggg aca tgg caa atg gac tgt aca cat tta gag gga aag gtt ata 2496
Leu Gly Thr Trp Gln Met Asp Cys Thr His Leu Glu Gly Lys Val Ile
820 825 830

ata gtg gca gtt cat gta gcc agt gga ttc ata gaa gca gaa gta ata 2544
Ile Val Ala Val His Val Ala Ser Gly Phe Ile Glu Ala Glu Val Ile

835 840 845

ccc caa gaa aca gga aga cag aca gct ctc ttc ctg tta aag ctg gcc 2592
Pro Gln Glu Thr Gly Arg Gln Thr Ala Leu Phe Leu Leu Lys Leu Ala
850 855 860

agc aga tgg cct atc aca cac ctg cac aca gac aac ggt gcc aac ttc 2640
Ser Arg Trp Pro Ile Thr His Leu His Thr Asp Asn Gly Ala Asn Phe
865 870 875 880

act tca caa gat gtg aaa atg gca gcc tgg tgg ata ggg ata gaa caa 2688
Thr Ser Gln Asp Val Lys Met Ala Ala Trp Trp Ile Gly Ile Glu Gln
885 890 895

aca ttc gga gtg ccc tat aat cca gaa agt cag gga gta gta gaa gca 2736
Thr Phe Gly Val Pro Tyr Asn Pro Glu Ser Gln Gly Val Val Glu Ala
900 905 910

atg aac cat cat ctg aaa aat cag ata gac aga att aga gat cag gca 2784
Met Asn His His Leu Lys Asn Gln Ile Asp Arg Ile Arg Asp Gln Ala
915 920 925

gta tca ata gag aca gtt gtg tta atg gca act cac tgc atg aat ttt 2832
Val Ser Ile Glu Thr Val Val Leu Met Ala Thr His Cys Met Asn Phe
930 935 940

aaa aga agg gga gga ata ggg gat atg acc cct gca gaa aga ata gtc 2880
Lys Arg Arg Gly Gly Ile Gly Asp Met Thr Pro Ala Glu Arg Ile Val
945 950 955 960

aac atg ata act aca gaa caa gaa ata caa ttc ctc caa aca aaa aat 2928
Asn Met Ile Thr Thr Glu Gln Glu Ile Gln Phe Leu Gln Thr Lys Asn
965 970 975

tta aaa ttc caa aat ttc cgg gtc tat tac aga gaa ggc aga gat caa 2976
Leu Lys Phe Gln Asn Phe Arg Val Tyr Tyr Arg Glu Gly Arg Asp Gln
980 985 990

ctc tgg aag gga cct ggt gat cta ttg tgg aaa ggg gaa gga gca gtc 3024
Leu Trp Lys Gly Pro Gly Asp Leu Leu Trp Lys Gly Glu Gly Ala Val
995 1000 1005

atc ata aag gta ggg aca gaa atc aaa gta ata ccc aga aga aaa gca 3072
Ile Ile Lys Val Gly Thr Glu Ile Lys Val Ile Pro Arg Arg Lys Ala

1010 1015 1020

aag atc ata agg aac tat gga gga gga aaa gaa ttg gat tgc agt gcc 3120
Lys Ile Ile Arg Asn Tyr Gly Gly Gly Lys Glu Leu Asp Cys Ser Ala
1025 1030 1035 1040

gac gtg gag gat acc atg cag gct aga gag gtg gca cag tct aat taa 3168
Asp Val Glu Asp Thr Met Gln Ala Arg Glu Val Ala Gln Ser Asn
1045 1050 1055

<210> 39
<211> 1055
<212> PRT
<213> Human immunodeficiency virus type 2

<400> 39
Met Leu Glu Met Trp Thr Ala Arg Thr His His Val Lys Met Pro Arg
1 5 10 15

Lys Thr Gly Gly Phe Phe Arg Val Arg Pro Leu Gly Lys Glu Ala Ser
20 25 30

Gln Phe Pro Arg Pro Gly Thr Pro Gly Asp Ser Ala Ile Cys Ala Pro
35 40 45

Asp Glu Pro Ser Ile Arg His Asp Thr Ser Gly Cys Asp Ser Ile Cys
50 55 60

Thr Pro Cys Arg Ser Ser Arg Gly Asp Ala Lys Glu Leu His Ala Thr
65 70 75 80

Arg Glu Glu Ala Glu Gly Glu Gln Arg Glu Thr Leu Gln Gly Gly Asp
85 90 95

Arg Gly Phe Ala Ala Pro Gln Phe Ser Leu Trp Arg Arg Pro Val Val
100 105 110

Lys Ala Thr Ile Glu Gly Gln Ser Val Glu Val Leu Leu Asp Thr Gly
115 120 125

Ala Asp Asp Ser Ile Val Ala Gly Ile Glu Leu Gly Ser Asn Tyr Thr
130 135 140

Pro Lys Ile Val Gly Gly Ile Gly Gly Phe Ile Asn Thr Asn Glu Tyr
145 150 155 160

Lys Asn Val Arg Ile Glu Val Val Gly Lys Arg Val Arg Ala Thr Val
165 170 175

Met Thr Gly Asp Thr Pro Ile Asn Ile Phe Gly Arg Asn Ile Leu Asn
180 185 190

Ser Leu Gly Met Thr Leu Asn Phe Pro Val Ala Arg Ile Glu Pro Val
195 200 205

Lys Val Gln Leu Lys Pro Glu Lys Asp Gly Pro Lys Ile Arg Gln Trp
210 215 220

Pro Leu Ser Lys Glu Lys Ile Leu Ala Leu Lys Glu Ile Cys Glu Lys
225 230 235 240

Met Glu Lys Glu Gly Gln Leu Glu Glu Ala Pro Pro Thr Asn Pro Tyr
245 250 255

Asn Ser Pro Thr Phe Ala Ile Lys Lys Lys Asp Lys Asn Lys Trp Arg
260 265 270

Met Leu Ile Asp Phe Arg Glu Leu Asn Lys Val Thr Gln Glu Phe Thr
275 280 285

Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Ala Ser Lys Lys
290 295 300

Arg Ile Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu
305 310 315 320

Asp Pro Asp Phe Arg Gln Tyr Thr Ala Phe Thr Leu Pro Ala Val Asn
325 330 335

Asn Ala Glu Pro Gly Lys Arg Tyr Leu Tyr Lys Val Leu Pro Gln Gly
340 345 350

Trp Lys Gly Ser Pro Ala Ile Phe Gln Tyr Thr Met Ala Lys Val Leu
355 360 365

Asp Pro Phe Arg Lys Ala Asn Asn Asp Val Thr Ile Ile Gln Tyr Met
370 375 380

Asp Asp Ile Leu Val Ala Ser Asp Arg Ser Asp Leu Glu His Asp Arg
385 390 395 400

Val Val Ser Gln Leu Lys Glu Leu Leu Asn Asn Met Gly Phe Ser Thr
405 410 415

Pro Glu Glu Lys Phe Gln Lys Asp Pro Pro Phe Lys Trp Met Gly Tyr
420 425 430

Glu Leu Trp Pro Lys Lys Trp Lys Leu Gln Lys Ile Gln Leu Pro Glu
435 440 445

Lys Glu Val Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Val Leu
450 455 460

Asn Trp Ala Ala Gln Leu Phe Pro Gly Ile Lys Thr Arg His Ile Cys
465 470 475 480

Lys Leu Ile Arg Gly Lys Met Thr Leu Thr Glu Glu Val Gln Trp Thr
485 490 495

Glu Leu Ala Glu Ala Glu Phe Gln Glu Asn Lys Ile Ile Leu Glu Gln
500 505 510

Glu Gln Glu Gly Ser Tyr Tyr Lys Glu Gly Val Pro Leu Glu Ala Thr
515 520 525

Val Gln Lys Asn Leu Ala Asn Gln Trp Thr Tyr Lys Ile His Gln Gly
530 535 540

Asp Lys Ile Leu Lys Val Gly Lys Tyr Ala Lys Val Lys Asn Thr His
545 550 555 560

Thr Asn Gly Val Arg Leu Leu Ala His Val Val Gln Lys Ile Gly Lys
565 570 575

Glu Ala Leu Val Ile Trp Gly Glu Ile Pro Met Phe His Leu Pro Val
580 585 590

Glu Arg Glu Thr Trp Asp Gln Trp Trp Thr Asp Tyr Trp Gln Val Thr
595 600 605

Trp Ile Pro Glu Trp Asp Phe Val Ser Thr Pro Pro Leu Ile Arg Leu
610 615 620

Ala Tyr Asn Leu Val Lys Asp Pro Leu Glu Gly Val Glu Thr Tyr Tyr
625 630 635 640

Thr Asp Gly Ser Cys Asn Lys Ala Ser Lys Glu Gly Lys Ala Gly Tyr
645 650 655

Val Thr Asp Arg Gly Lys Asp Lys Val Lys Pro Leu Glu Gln Thr Thr
660 665 670

Asn Gln Gln Ala Glu Leu Glu Ala Phe Ala Leu Ala Leu Gln Asp Ser
675 680 685

Gly Pro Gln Val Asn Ile Ile Val Asp Ser Gln Tyr Val Met Gly Ile
690 695 700

Val Ala Ala Gln Pro Thr Glu Thr Glu Ser Pro Ile Val Arg Glu Ile
705 710 715 720

Ile Glu Glu Met Ile Lys Lys Glu Lys Ile Tyr Val Gly Trp Val Pro
725 730 735

Ala His Lys Gly Leu Gly Gly Asn Gln Glu Val Asp His Leu Val Ser
740 745 750

Gln Gly Ile Arg Gln Ile Leu Phe Leu Glu Lys Ile Glu Pro Ala Gln
755 760 765

Glu Glu His Glu Lys Tyr His Asn Asn Val Lys Glu Leu Val His Lys
770 775 780

Phe Gly Ile Pro Gln Leu Val Ala Arg Gln Ile Val Asn Ser Cys Asp
785 790 795 800

Lys Cys Gln Gln Lys Gly Glu Ala Ile His Gly Gln Val Asn Ser Glu
805 810 815

Leu Gly Thr Trp Gln Met Asp Cys Thr His Leu Glu Gly Lys Val Ile
820 825 830

Ile Val Ala Val His Val Ala Ser Gly Phe Ile Glu Ala Glu Val Ile
835 840 845

Pro Gln Glu Thr Gly Arg Gln Thr Ala Leu Phe Leu Leu Lys Leu Ala
850 855 860

Ser Arg Trp Pro Ile Thr His Leu His Thr Asp Asn Gly Ala Asn Phe
865 870 875 880

Thr Ser Gln Asp Val Lys Met Ala Ala Trp Trp Ile Gly Ile Glu Gln
885 890 895

Thr Phe Gly Val Pro Tyr Asn Pro Glu Ser Gln Gly Val Val Glu Ala
900 905 910

Met Asn His His Leu Lys Asn Gln Ile Asp Arg Ile Arg Asp Gln Ala
915 920 925

Val Ser Ile Glu Thr Val Val Leu Met Ala Thr His Cys Met Asn Phe
930 935 940

Lys Arg Arg Gly Gly Ile Gly Asp Met Thr Pro Ala Glu Arg Ile Val
945 950 955 960

Asn Met Ile Thr Thr Glu Gln Glu Ile Gln Phe Leu Gln Thr Lys Asn
965 970 975

Leu Lys Phe Gln Asn Phe Arg Val Tyr Tyr Arg Glu Gly Arg Asp Gln
980 985 990

Leu Trp Lys Gly Pro Gly Asp Leu Leu Trp Lys Gly Glu Gly Ala Val
995 1000 1005

Ile Ile Lys Val Gly Thr Glu Ile Lys Val Ile Pro Arg Arg Lys Ala
1010 1015 1020

Lys Ile Ile Arg Asn Tyr Gly Gly Gly Lys Glu Leu Asp Cys Ser Ala
025 1030 1035 1040

Asp Val Glu Asp Thr Met Gln Ala Arg Glu Val Ala Gln Ser Asn
1045 1050 1055

<210> 40
<211> 3643
<212> DNA
<213> Murine leukemia virus

<220>
<221> CDS

<220> (1)...(3597)
<223> MMLV Pol

<400> 46

gga ggt cag ggt cag gag ccc ccc cct gaa ccc agg ata acc ctc aaa 48
Gly Gly Gln Gly Gln Glu Pro Pro Pro Glu Pro Arg Ile Thr Leu Lys
1 5 10 15

gtc gag gag caa ccc gtc acc ttc ctg gta gat act ggg gcc caa cac 96
Val Glu Glu Gln Pro Val Thr Phe Leu Val Asp Thr Gly Ala Gln His
20 25 30

tcc gtg ctg acc caa aat cct gga ccc cta agt gat aag tct gcc tgg 144
Ser Val Leu Thr Gln Asn Pro Gly Pro Leu Ser Asp Lys Ser Ala Trp
35 40 45

gtc caa gag gct act gga gga aag cgg tat cgc tgg acc acg gat cgc 192
Val Gln Gly Ala Thr Gly Gly Lys Arg Tyr Arg Trp Thr Thr Asp Arg
50 55 60

aaa gta cat cta gct acc ggt aag gtc acc cac tct ttc ctc cat gta 240
Lys Val His Leu Ala Thr Gly Lys Val Thr His Ser Phe Leu His Val
65 70 75 80

cca gac tgt ccc tat cct ctg tta gga aga gat ttg ctg act aaa cta 288
Pro Asp Cys Pro Tyr Pro Leu Leu Gly Arg Asp Leu Leu Thr Lys Leu
85 90 95

aaa gcc caa atc cac ttt gag gga tca gga gct cag gtt atg gga cca 336
Lys Ala Gln Ile His Phe Glu Gly Ser Gly Ala Gln Val Met Gly Pro
100 105 110

ata ggg cag ccc ctg caa gtg ttg acc cta aat ata gaa gat gag cat 384
Met Gly Gln Pro Leu Gln Val Leu Thr Leu Asn Ile Glu Asp Glu His
115 120 125

cgg cta cat gag acc tca aaa gag cca gat gtt tct cta ggg tcc aca 432
Arg Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr
130 135 140

tgg ctg tct gat ttt cct cag gcc tgg gcg gaa acc ggg ggc atg gga 480
Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met Gly
145 150 155 160

ctg gca gtt cgc caa gct cct ctg atc ata cct ctg aaa gca acc tct 528
Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala Thr Ser
165 170 175

acc ccc gtg tcc ata aaa caa tac ccc atg tca caa gaa gcc aga ctg 576
Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu Ala Arg Leu
180 185 190

ggg atc aag ccc cac ata cag aga ctg ttg gac cag gga ata ctg gta 624
Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln Gly Ile Leu Val
195 200 205

ccc tgc cag tcc ccc tgg aac acg ccc ctg cta ccc gtt aag aaa cca 672
Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu Pro Val Lys Lys Pro
210 215 220

ggg act aat gat tat agg cct gtc cag gat ctg aga gaa gtc aac aag 720
Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp Leu Arg Glu Val Asn Lys
225 230 235 240

cgg gtg gaa gac atc cac ccc acc gtg ccc aac cct tac aac ctc ttg 768
Arg Val Glu Asp Ile His Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu
245 250 255

agc ggg ctc cca ccg tcc cac cag tgg tac act gtg ctt gat tta aag 816
Ser Gly Leu Pro Pro Ser His Gln Trp Tyr Thr Val Leu Asp Leu Lys
260 265 270

gat gcc ttt ttc tgc ctg aga ctc cac ccc acc agt cag cct ctc ttc 864
Asp Ala Phe Phe Cys Leu Arg Leu His Pro Thr Ser Gln Pro Leu Phe
275 280 285

gcc ttt gag tgg aga gat cca gag atg gga atc tca gga caa ttg acc 912
Ala Phe Glu Trp Arg Asp Pro Glu Met Gly Ile Ser Gly Gln Leu Thr
290 295 300

tgg acc aga ctc cca cag ggt ttc aaa aac agt ccc acc ctg ttt gat 960
Trp Thr Arg Leu Pro Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp
305 310 315 320

gag gca ctg cac aga gac cta gca gac ttc cgg atc cag cac cca gac 1008
Glu Ala Leu His Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp
325 330 335

ttg atc ctg cta cag tac gtg gat gac tta ctg ctg gcc gcc act tct 1056
Leu Ile Leu Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser
340 345 350

gag cta gac tgc caa caa ggt act cgg gcc ctg tta caa acc cta ggg 1104
Glu Leu Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly
355 360 365

aac ctc ggg tat cgg gcc tcg gcc aag aaa gcc caa att tgc cag aaa 1152
Asn Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys
370 375 380

cag gtc aag tat ctg ggg tat ctt cta aaa gag ggt cag aga tgg ctg 1200
Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp Leu
385 390 395 400

act gag gcc aga aaa gag act gtg atg ggg cag cct act ccg aag acc 1248
Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro Lys Thr
405 410 415

cct cga caa cta agg gag ttc cta ggg acg gca ggc ttc tgt cgc ctc 1296
Pro Arg Gln Leu Arg Glu Phe Leu Gly Thr Ala Gly Phe Cys Arg Leu
420 425 430

tgg atc cct ggg ttt gca gaa atg gca gcc ccc ttg tac cct ctc acc 1344
Trp Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr
435 440 445

aaa acg ggg act ctg ttt aat tgg ggc cca gac caa caa aag gcc tat 1392
Lys Thr Gly Thr Leu Phe Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr
450 455 460

caa gaa atc aag caa gct ctt cta act gcc cca gcc ctg ggg ttg cca 1440
Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro
465 470 475 480

gat ttg act aag ccc ttt gaa ctc ttt gtc gac gag aag cag ggc tac 1488
Asp Leu Thr Lys Pro Phe Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr
485 490 495

gcc aaa ggt gtc cta acg caa aaa ctg gga cct tgg cgt cgg ccg gtg 1536
Ala Lys Gly Val Leu Thr Gln Lys Leu Gly Pro Trp Arg Arg Pro Val
500 505 510

gcc tac ctg tcc aaa aag cta gac cca gta gca gct ggg tgg ccc cct 1584
Ala Tyr Leu Ser Lys Lys Leu Asp Pro Val Ala Ala Gly Trp Pro Pro
515 520 525

tgc cta cgg atg gta gca gcc att gcc gta ctg aca aag gat gca ggc 1632
Cys Leu Arg Met Val Ala Ala Ile Ala Val Leu Thr Lys Asp Ala Gly
530 535 540

aag cta acc atg gga cag cca cta gtc att ctg gcc ccc cat gca gta 1680
Lys Leu Thr Met Gly Gln Pro Leu Val Ile Leu Ala Pro His Ala Val
545 550 555 560

gag gca cta gtc aaa caa ccc ccc gac cgc tgg ctt tcc aac gcc cgg 1728
Glu Ala Leu Val Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg
565 570 575

atg act cac tat cag gcc ttg ctt ttg gac acg gac cgg gtc cag ttc 1776
Met Thr His Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe
580 585 590

gga ccg gtg gta gcc ctg aac ccg gct acg ctg ctc cca ctg cct gag 1824
Gly Pro Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu
595 600 605

gaa ggg ctg caa cac aac tgc ctt gat atc ctg gcc gaa gcc cac gga 1872
Glu Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly
610 615 620

acc cga ccc gac cta acg gac cag ccg ctc cca gac gcc gac cac acc 1920
Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His Thr
625 630 635 640

tgg tac acg gat gga agc agt ctc tta caa gag gga cag cgt aag gcg 1968
Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg Lys Ala
645 650 655

gga gct gcg gtg acc acc gag acc gag gta atc tgg gct aaa gcc ctg 2016
Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala Lys Ala Leu
660 665 670

cca gcc ggg aca tcc gct cag cgg gct gaa ctg ata gca ctc acc cag 2064
Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln
675 680 685

gcc cta aag atg gca gaa ggt aag aag cta aat gtt tat act gat agc 2112
Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser
690 695 700

cgt tat gct ttt gct act gcc cat atc cat gga gaa ata tac aga agg 2160
Arg Tyr Ala Phe Ala Thr Ala His Ile His Gly Glu Ile Tyr Arg Arg
705 710 715 720

cgt ggg ttg ctc aca tca gaa ggc aaa gag atc aaa aat aaa gac gag 2208
Arg Gly Leu Leu Thr Ser Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu
725 730 735

atc ttg gcc cta cta aaa gcc ctc ttt ctg ccc aaa aga ctt agc ata 2256
Ile Leu Ala Leu Leu Lys Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile
740 745 750

atc cat tgt cca gga cat caa aag gga cac agc gcc gag gct aga ggc 2304
Ile His Cys Pro Gly His Gln Lys Gly His Ser Ala Glu Ala Arg Gly
755 760 765

aac cgg atg gct gac caa gcg gcc cga aag gca gcc atc aca gag act 2352
Asn Arg Met Ala Asp Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr
770 775 780

cca gac acc tct acc ctc ctc ata gaa aat tca tca ccc tac acc tca 2400
Pro Asp Thr Ser Thr Leu Leu Ile Glu Asn Ser Ser Pro Tyr Thr Ser
785 790 795 800

gaa cat ttt cat tac aca gtg act gat ata aag gac cta acc aag ttg 2448
Glu His Phe His Tyr Thr Val Thr Asp Ile Lys Asp Leu Thr Lys Leu
805 810 815

ggg gcc att tat gat aaa aca aag aag tat tgg gtc tac caa gga aaa 2496
Gly Ala Ile Tyr Asp Lys Thr Lys Lys Tyr Trp Val Tyr Gln Gly Lys
820 825 830

cct gtg atg cct gac cag ttt act ttt gaa tta tta gac ttt ctt cat 2544
Pro Val Met Pro Asp Gln Phe Thr Phe Glu Leu Leu Asp Phe Leu His
835 840 845

cag ctg act cac ctc agc ttc tca aaa atg aag gct ctc cta gag aga 2592
Gln Leu Thr His Leu Ser Phe Ser Lys Met Lys Ala Leu Leu Glu Arg
850 855 860

agc cac agt ccc tac tac atg ctg aac cgg gat cga aca ctc aaa aat 2640
Ser His Ser Pro Tyr Tyr Met Leu Asn Arg Asp Arg Thr Leu Lys Asn
865 870 875 880

atc act gag acc tgc aaa gct tgt gca caa gtc aac gcc agc aag tct 2688
Ile Thr Glu Thr Cys Lys Ala Cys Ala Gln Val Asn Ala Ser Lys Ser
885 890 895

gcc gtt aaa cag gga act agg gtc cgc ggg cat cgg ccc ggc act cat 2736
Ala Val Lys Gln Gly Thr Arg Val Arg Gly His Arg Pro Gly Thr His
900 905 910

tgg gag atc gat ttc acc gag ata aag ccc gga ttg tat ggc tat aaa 2784
Trp Glu Ile Asp Phe Thr Glu Ile Lys Pro Gly Leu Tyr Gly Tyr Lys
915 920 925

tat ctt cta gtt ttt ata gat acc ttt tct ggc tgg ata gaa gcc ttc 2832
Tyr Leu Leu Val Phe Ile Asp Thr Phe Ser Gly Trp Ile Glu Ala Phe
930 935 940

cca acc aag aaa gaa acc gcc aag gtc gta acc aag aag cta cta gag 2880
Pro Thr Lys Lys Glu Thr Ala Lys Val Val Thr Lys Lys Leu Leu Glu
945 950 955 960

gag atc ttc ccc agg ttc ggc atg cct cag gta ttg gga act gac aat 2928
Glu Ile Phe Pro Arg Phe Gly Met Pro Gln Val Leu Gly Thr Asp Asn
965 970 975

ggg cct gcc ttc gtc tcc aag gtg agt cag aca gtg gcc gat ctg ttg 2976
Gly Pro Ala Phe Val Ser Lys Val Ser Gln Thr Val Ala Asp Leu Leu
980 985 990

ggg att gat tgg aaa tta cat tgt gca tac aga ccc caa agc tca ggc 3024
Gly Ile Asp Trp Lys Leu His Cys Ala Tyr Arg Pro Gln Ser Ser Gly
995 1000 1005

cag gta gaa aga atg aat aga acc atc aag gag act tta act aaa tta 3072
Gln Val Glu Arg Met Asn Arg Thr Ile Lys Glu Thr Leu Thr Lys Leu
1010 1015 1020

acg ctt gca act ggc tct aga gac tgg gtg ctc cta ctc ccc tta gcc 3120
Thr Leu Ala Thr Gly Ser Arg Asp Trp Val Leu Leu Leu Pro Leu Ala
1025 1030 1035 1040

ctg tac cga gcc cgc aac acg ccg ggc ccc cat ggc ctc acc cca tat 3168
Leu Tyr Arg Ala Arg Asn Thr Pro Gly Pro His Gly Leu Thr Pro Tyr
1045 1050 1055

gag atc tta tat ggg gca ccc ccg ccc ctt gta aac ttc cct gac cct 3216
Glu Ile Leu Tyr Gly Ala Pro Pro Pro Leu Val Asn Phe Pro Asp Pro
1060 1065 1070

gac atg aca aga gtt act aac agc ccc tct ctc caa gct cac tta cag 3264
Asp Met Thr Arg Val Thr Asn Ser Pro Ser Leu Gln Ala His Leu Gln
1075 1080 1085

gct ctc tac tta gtc cag cac gaa gtc tgg aga cct ctg gcg gca gcc 3312
Ala Leu Tyr Leu Val Gln His Glu Val Trp Arg Pro Leu Ala Ala Ala
1090 1095 1100

tac caa gaa caa ctg gac cga ccg gtg gta cct cac cct tac cga gtc 3360
Tyr Gln Glu Gln Leu Asp Arg Pro Val Val Pro His Pro Tyr Arg Val
1105 1110 1115 1120

ggc gac aca gtg tgg gtc cgc cga cac cag act aag aac cta gaa cct 3408
Gly Asp Thr Val Trp Val Arg Arg His Gln Thr Lys Asn Leu Glu Pro
1125 1130 1135

cgc tgg aaa gga cct tac aca gtc ctg ctg acc acc ccc acc gcc ctc 3456
Arg Trp Lys Gly Pro Tyr Thr Val Leu Leu Thr Thr Pro Thr Ala Leu
1140 1145 1150

aaa gta gac ggc atc gca gct tgg ata cac gcc gcc cac gtg aag gct 3504
Lys Val Asp Gly Ile Ala Ala Trp Ile His Ala Ala His Val Lys Ala
1155 1160 1165

gcc gac ccc ggg ggt gga cca tcc tct aga ctg aca tgg cgc gtt caa 3552
Ala Asp Pro Gly Gly Gly Pro Ser Ser Arg Leu Thr Trp Arg Val Gln
1170 1175 1180

cgc tct caa aac ccc tta aaa ata agg tta acc cgc gag gcc ccc 3597
Arg Ser Gln Asn Pro Leu Lys Ile Arg Leu Thr Arg Glu Ala Pro
1185 1190 1195

taatccccct aatttcttctg atgctcagag gggtcagtac tgcttc 3643

<210> 41

<211> 1199
<212> PRT
<213> Murine leukemia virus

<400> 41

Gly Gly Gln Gly Gln Glu Pro Pro Pro Glu Pro Arg Ile Thr Leu Lys
1 5 10 15

Val Gly Gly Gln Pro Val Thr Phe Leu Val Asp Thr Gly Ala Gln His
20 25 30

Ser Val Leu Thr Gln Asn Pro Gly Pro Leu Ser Asp Lys Ser Ala Trp
35 40 45

Val Gln Gly Ala Thr Gly Gly Lys Arg Tyr Arg Trp Thr Thr Asp Arg
50 55 60

Lys Val His Leu Ala Thr Gly Lys Val Thr His Ser Phe Leu His Val
65 70 75 80

Pro Asp Cys Pro Tyr Pro Leu Leu Gly Arg Asp Leu Leu Thr Lys Leu
85 90 95

Lys Ala Gln Ile His Phe Glu Gly Ser Gly Ala Gln Val Met Gly Pro
100 105 110

Met Gly Gln Pro Leu Gln Val Leu Thr Leu Asn Ile Glu Asp Glu His
115 120 125

Arg Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr
130 135 140

Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met Gly
145 150 155 160

Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala Thr Ser
165 170 175

Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu Ala Arg Leu
180 185 190

Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln Gly Ile Leu Val
195 200 205

Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu Pro Val Lys Lys Pro
210 215 220

Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp Leu Arg Glu Val Asn Lys
225 230 235 240

Arg Val Glu Asp Ile His Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu
245 250 255

Ser Gly Leu Pro Pro Ser His Gln Trp Tyr Thr Val Leu Asp Leu Lys
260 265 270

Asp Ala Phe Phe Cys Leu Arg Leu His Pro Thr Ser Gln Pro Leu Phe
275 280 285

Ala Phe Glu Trp Arg Asp Pro Glu Met Gly Ile Ser Gly Gln Leu Thr
290 295 300

Trp Thr Arg Leu Pro Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp
305 310 315 320

Glu Ala Leu His Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp
325 330 335

Leu Ile Leu Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser
340 345 350

Glu Leu Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly
355 360 365

Asn Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys
370 375 380

Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp Leu
385 390 395 400

Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro Lys Thr
405 410 415

Pro Arg Gln Leu Arg Glu Phe Leu Gly Thr Ala Gly Phe Cys Arg Leu
420 425 430

Trp Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr
435 440 445

Lys Thr Gly Thr Leu Phe Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr
450 455 460

Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro
465 470 475 480

Asp Leu Thr Lys Pro Phe Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr
485 490 495

Ala Lys Gly Val Leu Thr Gln Lys Leu Gly Pro Trp Arg Arg Pro Val
500 505 510

Ala Tyr Leu Ser Lys Lys Leu Asp Pro Val Ala Ala Gly Trp Pro Pro
515 520 525

Cys Leu Arg Met Val Ala Ala Ile Ala Val Leu Thr Lys Asp Ala Gly
530 535 540

Lys Leu Thr Met Gly Gln Pro Leu Val Ile Leu Ala Pro His Ala Val
545 550 555 560

Glu Ala Leu Val Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg
565 570 575

Met Thr His Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe
580 585 590

Gly Pro Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu
595 600 605

Glu Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly
610 615 620

Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His Thr
625 630 635 640

Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg Lys Ala
645 650 655

Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala Lys Ala Leu
660 665 670

Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln
675 680 685

Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser
690 695 700

Arg Tyr Ala Phe Ala Thr Ala His Ile His Gly Glu Ile Tyr Arg Arg
705 710 715 720

Arg Gly Leu Leu Thr Ser Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu
725 730 735

Ile Leu Ala Leu Leu Lys Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile
740 745 750

Ile His Cys Pro Gly His Gln Lys Gly His Ser Ala Glu Ala Arg Gly
755 760 765

Asn Arg Met Ala Asp Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr
770 775 780

Pro Asp Thr Ser Thr Leu Leu Ile Glu Asn Ser Ser Pro Tyr Thr Ser
785 790 795 800

Glu His Phe His Tyr Thr Val Thr Asp Ile Lys Asp Leu Thr Lys Leu
805 810 815

Gly Ala Ile Tyr Asp Lys Thr Lys Lys Tyr Trp Val Tyr Gln Gly Lys
820 825 830

Pro Val Met Pro Asp Gln Phe Thr Phe Glu Leu Leu Asp Phe Leu His
835 840 845

Gln Leu Thr His Leu Ser Phe Ser Lys Met Lys Ala Leu Leu Glu Arg
850 855 860

Ser His Ser Pro Tyr Tyr Met Leu Asn Arg Asn Arg Thr Leu Lys Asn
865 870 875 880

Ile Thr Glu Thr Cys Lys Ala Cys Ala Gln Val Asn Ala Ser Lys Ser
885 890 895

Ala Val Lys Gln Gly Thr Arg Val Arg Gly His Arg Pro Gly Thr His
900 905 910

Trp Glu Ile Asp Phe Thr Glu Ile Lys Pro Gly Leu Tyr Gly Tyr Lys
915 920 925

Tyr Leu Leu Val Phe Ile Asp Thr Phe Ser Gly Trp Ile Glu Ala Phe
930 935 940

Pro Thr Lys Arg Glu Thr Ala Lys Val Val Thr Lys Lys Leu Leu Glu
945 950 955 960

Glu Ile Phe Pro Arg Phe Gly Met Pro Gln Val Leu Gly Thr Asp Asn
965 970 975

Gly Pro Ala Phe Val Ser Lys Val Ser Gln Thr Val Ala Asp Leu Leu
980 985 990

Gly Ile Asp Trp Lys Leu His Cys Ala Tyr Arg Pro Gln Ser Ser Gly
995 1000 1005

Gln Val Glu Arg Met Asn Arg Thr Ile Lys Glu Thr Leu Thr Lys Leu
1010 1015 1020

Thr Leu Ala Thr Gly Ser Arg Asp Trp Val Leu Leu Leu Pro Leu Ala
1025 1030 1035 1040

Leu Tyr Arg Ala Arg Asn Thr Pro Gly Pro His Gly Leu Thr Pro Tyr
1045 1050 1055

Glu Ile Leu Tyr Gly Ala Pro Pro Pro Leu Val Asn Phe Pro Asp Pro
1060 1065 1070

Asp Met Thr Arg Val Thr Asn Ser Pro Ser Leu Gln Ala His Leu Gln
1075 1080 1085

Ala Leu Tyr Leu Val Gln His Glu Val Trp Arg Pro Leu Ala Ala Ala
1090 1095 1100

Tyr Gln Glu Gln Leu Asp Arg Pro Val Val Pro His Pro Tyr Arg Val
1105 1110 1115 1120

Gly Asp Thr Val Trp Val Arg Arg His Gln Thr Lys Asn Leu Glu Pro
1125 1130 1135

Arg Trp Lys Gly Pro Tyr Thr Val Leu Leu Thr Thr Pro Thr Ala Leu
1140 1145 1150

Lys Val Asp Gly Ile Ala Ala Trp Ile His Ala Ala His Val Lys Ala
1155 1160 1165

Ala Asp Pro Gly Gly Gly Pro Ser Ser Arg Leu Thr Trp Arg Val Gln
1170 1175 1180

Arg Gln Asn Pro Leu Lys Ile Arg Leu Thr Arg Glu Ala Pro
1185 1190 1195

<42>
<211> 2709
<212> DNA
<213> Human immunodeficiency virus type 1

<220>
<221> CDS
<222> (1)..(2709)

<400> 42
atg ata ggg gga att gga ggt ttt atc aaa gta aga cag tat gat cag 48
Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln
1 5 10 15

ata ctc ata gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta 96
Ile Leu Ile Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val
20 25 30

gga cct aca cct gtc aac ata att gga aga aat ctg ttg act cag att 144
Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile
35 40 45

ggt tgc act tta aat ttt ccc att agt cct att gaa act gta cca gta 192
Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val
50 55 60

aaa tta aag cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg 240
Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu
65 70 75 80

aca gaa gaa aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa 288
Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu
85 90 95

aag gaa ggg aaa att tca aaa att ggg cct gaa aat cca tac aat act 336
Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr
100 105 110

```
cca gta ttt gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta   384
Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu
        115                 120                 125

gta gat ttc aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt   432
Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val
    130                 135                 140

caa tta gga ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta   480
Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val
145                 150                 155                 160

aca gta ctg gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa   528
Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu
                165                 170                 175

gac ttc agg aag tat act gca ttt acc ata cct agt ata aac aat gag   576
Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu
            180                 185                 190

aca cca ggg att aga tat cag tac aat gtg ctt cca cag gga tgg aaa   624
Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys
        195                 200                 205

gga tca cca gca ata ttc caa agt agc atg aca aaa atc tta gag cct   672
Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro
    210                 215                 220

ttt aga aaa caa aat cca gac ata gtt atc tat caa tac atg gat gat   720
Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp
225                 230                 235                 240

ttg tat gta gga tct gac tta gaa ata ggg cag cat aga aca aaa ata   768
Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile
                245                 250                 255

gag gag ctg aga caa cat ctg ttg agg tgg gga ctt acc aca cca gac   816
Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp
            260                 265                 270

aaa aaa cat cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc   864
Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu
        275                 280                 285
```

912
cat cct gat aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac
His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp
290 295 300

960
agc tgg act gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg
Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp
305 310 315 320

1008
gca agt cag att tac cca ggg att aaa gta agg caa tta tgt aaa ctc
Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu
325 330 335

1056
ctt aga gga acc aaa gca cta aca gaa gta ata cca cta aca gaa gaa
Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu
340 345 350

1104
gca gag cta gaa ctg gca gaa aac aga gag att cta aaa gaa cca gta
Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val
355 360 365

1152
cat gga gtg tat tat gac cca tca aaa gac tta ata gca gaa ata cag
His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln
370 375 380

1200
aag cag ggg caa ggc caa tgg aca tat caa att tat caa gag cca ttt
Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe
385 390 395 400

1248
aaa aat ctg aaa aca gga aaa tat gca aga acg agg ggt gcc cac act
Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Thr Arg Gly Ala His Thr
405 410 415

1296
aat gat gta aaa caa tta aca gag gca gtg caa aaa ata acc aca gaa
Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu
420 425 430

1344
agc ata gta ata tgg gga aag act cct aaa ttt aaa cta ccc ata caa
Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln
435 440 445

1392
aag gaa aca tgg gaa aca tgg tgg aca gag tat tgg caa gcc acc tgg
Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp
450 455 460

att cct gaa tgg gag ttt gtc aat acc cct cct tta gtg aaa tta tgg 1440
Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp
465 470 475 480

tat cag tta gag aaa gaa ccc ata gta gga gca gaa acg ttc tat gta 1488
Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val
485 490 495

gat ggg gca gct agc agg gag act aaa tta gga aaa gca gga tat gtt 1536
Asp Gly Ala Ala Ser Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val
500 505 510

act aat aga gga aga caa aaa gtt gtc acc cta act gac aca aca aat 1584
Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn
515 520 525

cag aag act gag tta caa gca att cat cta gct ttg cag gat tcg gga 1632
Gln Lys Thr Glu Leu Gln Ala Ile His Leu Ala Leu Gln Asp Ser Gly
530 535 540

tta gaa gta aat ata gta aca gac tca caa tat gca tta gga atc att 1680
Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile
545 550 555 560

caa gca caa cca gat aaa agt gaa tca gag tta gtc aat caa ata ata 1728
Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile Ile
565 570 575

gag cag tta ata aaa aag gaa aag gtc tat ctg gca tgg gta cca gca 1776
Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala
580 585 590

cac aaa gga att gga gga aat gaa caa gta gat aaa tta gtc agt gct 1824
His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala
595 600 605

gga atc agg aaa gta cta ttt tta gat gga ata gat aag gcc caa gat 1872
Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp
610 615 620

gaa cat gag aaa tat cac agt aat tgg aga gca atg gct agt gat ttt 1920
Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe
625 630 635 640

aac ctg cca cct gta gta gca aaa gaa ata gta gcc agc tgt gat aaa 1968
Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys
645 650 655

tgt cag cta aaa gga gaa gcc atg cat gga caa gta gac tgt agt cca 2016
Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro
660 665 670

gga ata tgg caa cta gat tgt aca cat tta gaa gga aaa gtt atc ctg 2064
Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu
675 680 685

gta gca gtt cat gta gcc agt gga tat ata gaa gca gaa gtt att cca 2112
Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro
690 695 700

gca gaa aca ggg cag gaa aca gca tac ttt ctt tta aaa tta gca gga 2160
Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly
705 710 715 720

aga tgg cca gta aaa aca ata cat aca gac aat ggc agc aat ttc acc 2208
Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr
725 730 735

agt act acg gtt aag gcc gcc tgt tgg tgg gcg gga atc aag cag gaa 2256
Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu
740 745 750

ttt gga att ccc tac aat ccc caa agt caa gga gta gta gaa tct atg 2304
Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met
755 760 765

aat aaa gaa tta aag aaa att ata ggc cag gta aga gat cag gct gaa 2352
Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu
770 775 780

cat ctt aag aca gca gta caa atg gca gta ttc atc cac aat ttt aaa 2400
His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys
785 790 795 800

aga aaa ggg ggg att ggg ggg tac agt gca ggg gaa aga ata gta gac 2448
Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp
805 810 815

ata ata gca aca gac ata caa act aaa gaa tta caa aaa caa att aca 2496
Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr
820 825 830

aaa att caa aat ttt cgg gtt tat tac agg gac agc aga gat cca ctt 2544
Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Leu
835 840 845

tgg aaa gga cca gca aag ctc ctc tgg aaa ggt gaa ggg gca gta gta 2592
Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val
850 855 860

ata caa gat aat agt gac ata aaa gta gtg cca aga aga aaa gca aag 2640
Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys
865 870 875 880

atc att agg gat tat gga aaa cag atg gca ggt gat gat tgt gtg gca 2688
Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala
885 890 895

agt aga cag gat gag gat tag 2709
Ser Arg Gln Asp Glu Asp
900

<210> 43
<211> 902
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 43
Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln
1 5 10 15

Ile Leu Ile Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val
20 25 30

Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile
35 40 45

Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val
50 55 60

Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu

65 70 75 80

Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu
85 90 95

Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr
100 105 110

Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu
115 120 125

Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val
130 135 140

Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val
145 150 155 160

Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu
165 170 175

Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu
180 185 190

Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys
195 200 205

Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro
210 215 220

Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp
225 230 235 240

Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile
245 250 255

Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp
260 265 270

Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu
275 280 285

His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp
290 295 300

Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp
305 310 315 320

Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu
325 330 335

Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu
340 345 350

Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val
355 360 365

His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln
370 375 380

Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe
385 390 395 400

Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Thr Arg Gly Ala His Thr
405 410 415

Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu
420 425 430

Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln
435 440 445

Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp
450 455 460

Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp
465 470 475 480

Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val
485 490 495

Asp Gly Ala Ala Ser Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val
500 505 510

Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn
515 520 525

Gln Lys Thr Glu Leu Gln Ala Ile His Leu Ala Leu Gln Asp Ser Gly
530 535 540

Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile
545 550 555 560

Gln Ala Gln Ile Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile Ile
565 570 575

Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala
580 585 590

His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala
595 600 605

Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp
610 615 620

Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe
625 630 635 640

Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys
645 650 655

Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro
660 665 670

Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu
675 680 685

Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro
690 695 700

Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly
705 710 715 720

Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr
725 730 735

Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu
740 745 750

Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met
755 760 765

Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu
770 775 780

His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys
785 790 795 800

Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp
805 810 815

Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr
820 825 830

Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Leu
835 840 845

Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val
850 855 860

Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys
865 870 875 880

Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala
885 890 895

Ser Arg Gln Asp Glu Asp
900

<210> 44
<400> 44
000

<210> 45
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BAhisXhol extend

<400> 45
accccgatca atccgctcga gttagtggtg gtggtggtgg tgtttagcct ctctcaaggg 60

at 62

<210> 46
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA His 10

<400> 46
accccgatca atccgctcga gttagtggtg gtggtggtgg tggtggtggt ggtgtttagc 60
ctctctcaag ggat 74

<210> 47
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA His 12

<400> 47
accccgatca atccgctcga gttagtggtg gtggtggtgg tggtggtggt ggtggtggtg 60
tttagcctct ctcaagggat 80

<210> 48
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Leu

<400> 48
accccgatca atccgctcga gttagaggag tagtaggagt agtttagcct ctctcaaggg 60
a 61

<210> 49
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Lys

<400> 49
acccgatca atccgctcga gttacttctt cttcttcttc tttttagcct ctctcaaggg 60

a 61

<210> 50
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Arg6

<400> 50
acccgatca atccgctcga gttaacgacg acgacgacga cgtttagcct ctctcaaggg 60

at 62

<210> 51
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Arg3X4

<400> 51
acccgatca atccgctcga gttaacgatt acgattacgc tgatatttag cctctctcaa 60

gggat 65

<210> 52
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Asp6

<400> 52
acccgatca atccgctcga gttaatcgtc gtcatcgtca tctttagcct ctctcaaggg 60

61

a

<210> 53
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Asp4

<400> 53
acccgatca atccgctcga gttaatcgtc atcgtcttta gcctctctca aggga 55

<210> 54
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Asp5

<400> 54
acccgatca atccgctcga gttaatcgtc atcgtcatct ttagcctctc tcaaggga 58

<210> 55
<211> 67
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Asp8

<400> 55
acccccgatca atccgctcga gttaatcgtc gtcatcgtca tcgtcatctt tagcctctct 60

caaggga 67

<210> 56
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Asp12

<400> 56
acccccgatca atccgctcga gttaatcgtc gtcatcgtca tcgtcatcgt catcgtcatc 60

tttagcctct ctcaaggga 79

<210> 57
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Glu6 XhoI

<400> 57
acccccgatca atccgctcga gttattcctc ttcctcttcc tctttagcct ctctcaaggg 60

a 61

<210> 58
<211> 79

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Glu12

<400> 58
accccgatca atccgctcga gttattcctc ttcctcttcc tcttcctctt cctcttcctc 60

tttagcctct ctcaaggga 79

<210> 59
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA XhoI

<400> 59
accccgatca atccgctcga gttatttagc ctctctcaag ggat 44

<210> 60
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 6Z0 His

<400> 60
accccgatca atccgctcga gttagtggtg gtggtggtgg tgggcctcca acgcaggggc 60

tga 63

<210> 61
<211> 64
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA LZIP 2 XhoI

<400> 61
acccgatca atccgatcga gttacagttt ctgtgtttca gcctgtttag cctctctcaa 60

64

<210> 62
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA LZIP 3 XhoI

<400> 62
acccgatca atccgatcga gttaaagcag gtcaatctca gagatcagtt tctgtgtttc 60

85

agcctgttta gcctctctca aggga

<210> 63
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA LZIP 4 XhoI

<400> 63
acccgatca atccgatcga gttacagctg ctcgttctgt ttacgaagca ggtcaatctc 60

66

agagat

<210> 64
<211> 87

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA LZIP 5 XhoI

<400> 64
accccgatca atccgctcga gttacagctg ttccagcttg tgtttcagct gctcgttctg 60

tttacgaagc aggtcaatct cagagat 87

<210> 65
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Cyst 2

<400> 65
accccgatca atccgctcga gttagcaaca tttagcctct ctcaaggga 49

<210> 66
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA Cyst 6

<400> 66
accccgatca atccgctcga gttagcaaca gcaacagcaa catttagcct ctctcaaggg 60

a 61

<210> 67
<211> 55
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA GPRP

<400> 67
accccgatca atccgctcga gttaaggacg aggacctta gcctctctca aggga 55

<210> 68
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA PRPG

<400> 68
accccgatca atccgctcga gttaaccagg acgaggttta gcctctctca aggga 55

<210> 69
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA WH

<400> 69
ccgatcaatc cgctcgagtt atgcacgagg atgtgcagct tccatgcac gagctttagc 60
ctctctcaag gga 73

<210> 70
<211> 70
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic DNA
- RM1BA 3PPG XhoI

<400> 70
accccgatca atccgctcga gttaaccagg aggaccagga ggaccaggag gtttagcctc 60

tctcaaggga 70

<210> 71
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA TRP

<400> 71
accccgatca atccgctcga gttaccacca ccaccaccac catttagcct ctctcaaggg 60

a 61

<210> 72
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- FM1BA Nhis SmaI

<400> 72
ataagggcca ctgttctccc cgggatgcac caccaccacc accacactgt tgcgctacat 60

ctggct 66

<210> 73
<211> 75
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA NWH SmaI

<400> 73
agggccactg ttctccccgg gatggctcgt gcatgggaag ctgcacatgc tcgtgcaact 60
gttgcgctac atctg 75

<210> 74
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 620

<400> 74
acccgatca atccgctcga gggcctccaa cgcaggggct ga 42

<210> 75
<400> 75
000

<210> 76
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 640 XhoI

<400> 76
gcaatgtatg ccctcaatct cgagaagtgt aaagtctgtc tgcca 45

<210> 77
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 660 XhoI

<400> 77
gcaatgtatg ccctcaatct cgagtatcgc cgatgaggcg gtatcca 47

<210> 78
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 680 XhoI

<400> 78
gcaatgtatg ccctcaatct cgagagccgt ggcccaatga tgtt 44

<210> 79
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 760 XhoI

<400> 79
gcaatgtatg ccctcaatct cgagcagttc cccctgtttg ctggt 45

<210> 80
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 800 XhoI

<400> 80

gcaatgtatg ccctcaatct cgagtcgtat tttaaccggg ggtcct 46

<210> 81
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 640 His XhoI

<400> 81
gcaatgtatg ccctcaatct cgagttaatg gtgatggtga tggtgaagtg taaagtctgt 60

ctgcca 66

<210> 82
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 660 His XhoI

<400> 82
gcaatgtatg ccctcaatct cgagttaatg gtgatggtga tggtgtatcg ccgatgaggc 60

ggtatcca 68

<210> 83
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 680 His XhoI

<400> 83
gcaatgtatg ccctcaatct cgagttaatg gtgatggtga tggtgagccg tggcccaatg 60

65

atgtt

<210> 84
<211> 67
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 800 His XhoI

<400> 84
gcaatgtatg ccctcaatct cgagttaatg gtgatggtga tggtgtcgta ttttaaccgg 60

67

gggtcct

<210> 85
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- F Cint XhoI

<400> 85
gcaatgtatg ccctcaatct cgagcacttt gagcgtggtg aaaac 45

<210> 86
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- R Cint SalI

<400> 86
gcctgcaaaa agagggctcg cgtcgaccgc ctcatctttc ttagtcacct c 51

<210> 87
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- KCint His SalI

<400> 87
gcctgcaaaa agaagctcg cgtcgactta gtggtggtgg tggtggtgcg cctcatcttt 60

cttagtcacc tt 72

<210> 88
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- F Cint 731 SalI

<400> 88
gcaatgtatg ccctcaatgt cgacgccaac cggctcctga aagat 45

<210> 89
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- FCint751 SalI

<400> 89
gcaatgtatg ccctcaatgt cgacatgaaa agaatcccca ccagc 45

<210> 90
<211> 51
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RCint 830 XhoI

<400> 90
qcctgcaaaa agagggctcg cctcgagtta cttatcagtg tccctgtttt t 51

<210> 91
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RCint 830 his XhoI

<400> 91
qcctgcaaaa agagggctcg cctcgagtta atggtgatgg tgatggtgct tatcagtgtc 60
cctgttttt 69

<210> 92
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
FDNPCR1 (D450N)

<400> 92
tgccgggacc cactgtcttt actaacgcct cctcaagcac ccataa 46

<210> 93
<211> 46
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic DNA
- RDNPCR1 (D450N)

<400> 93
ttatgggtgc ttgaggaggc gttagtaaag acagtgggtc ccggca 46

<210> 94
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- FDNPCR2 (D505N)

<400> 94
acgcccacta atgtagtgac taactccgcg tttgttgcga aaatgt 46

<210> 95
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RDNPCR2 (D505N)

<400> 95
acattttcgc aacaaacgcg gagttagtca ctacattagt gggcgt 46

<210> 96
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- FM1BA E484Q

<400> 96
ttgggggcaa gtgtacaaca actgcaagca cgcgctgtgg ccatggca 48

<210> 97
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA E484Q

<400> 97
tgccatggcc acagcgcgtg cttgcagttg ttgtacactt gccccaa 48

<210> 98
<211> 102
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BA NS1 XhoI

<400> 98
acgatcaatc cgctcgagtt attttgtttg gatacgtcca cgttttgttt cttgtgcggt 60
tgttacggtt gtttctactg ttttagcctc tctcaaggga ta 102

<210> 99
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- F Cint 771 SalI

<400> 99
gcaatgtatg ccctcaatgt cgacgagcgt ggtgaaaaca caaaa 45

<210> 100
<211> 66

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic DNA
- RM1BK 760 His XhoI

<400> 100
gcaatgtatg ccctcaatct cgagttaatg gtgatggtga tggtgcagtt ccccctgttt 60

gctggt 66

INTERNATIONAL SEARCH REPORT

National Application No
PCT/US 00/00896

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/54 C12N9/12 C12N15/70 C12N15/85 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	QUILLENT, CAROLINE ET AL: "Extensive regions of pol are required for efficient human immunodeficienc virus polyprotein processing and particle maturation" VIROLOGY (1996), 219(1), 29-36 ,1996, XP000608777 abstract page 30, right-hand column, last paragraph -page 36, left-hand column, paragraph 1 --- -/--	1-33

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 May 2000	24/05/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Montero Lopez, B

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NAOKO TANESE ET AL.: "Structural requirements for bacterial expression of stable, enzymatically active fusion proteins containing the human immunodeficiency virus reverse transcriptase" DNA, vol. 7, no. 6, 1988, pages 407-416, XP002137205 page 408, right-hand column, last paragraph -page 409, right-hand column, paragraph 1 page 410, left-hand column, paragraph 2 -right-hand column, paragraph 1 page 411, left-hand column, paragraph 2 -page 414, right-hand column, paragraph 2 ---	1-33
A	US 5 668 005 A (MICHAEL LESLIE KOTEWICZ ET AL.) 16 September 1997 (1997-09-16) cited in the application column 2, line 51 -column 8, line 39 column 10, line 56 -column 12, line 24 column 16, line 31 -column 18, line 2 ---	1-33
A	JP 07 039378 A (TAKARA SHUZO CO. LTD.) 10 February 1995 (1995-02-10) page 6 -page 11 -----	1-33

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/00896

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5668005 A	16-09-1997	US 5405776 A	11-04-1995
		US 5244797 A	14-09-1993
JP 7039378 A	10-02-1995	NONE	